Phenotype – genotype associations in a large cohort of patients with pulmonary arterial hypertension

Charaka Mayura Bandara Hadinnapola

Selwyn College

September 2018

This dissertation is submitted for the degree of

Doctor of Philosophy

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text.

It does not exceed the prescribed word limit for the Degree Committee for the Faculties of Clinical Medicine and Veterinary Medicine.

Summary: Phenotype – genotype associations in a large cohort of patients with pulmonary arterial hypertension. Charaka Mayura Bandara Hadinnapola

Idiopathic and heritable pulmonary arterial hypertension (PAH) are rare diseases with a poor prognosis. There is significant heterogeneity in clinical features at diagnosis, and in a proportion of patients there is a genetic cause of the disease. This clinical and genetic heterogeneity has hindered accurate risk stratification and the development of personalised treatments.

The National Institute of Health Research BioResource – Rare Diseases PAH Study and the Medical Research Council / British Heart Foundation National Cohort Study of Idiopathic and Heritable PAH were established to investigate disease pathogenesis through whole genome sequencing and deep phenotyping. One thousand and seventy patients were sequenced and 391 clinical variables were captured for each patient.

In patients with idiopathic PAH the age at diagnosis, gender, right atrial pressure and the transfer coefficient for carbon monoxide were identified as independent prognostic variables. However, the presence of rare and predicted deleterious variants in *BMPR2* was not of prognostic significance. Variants in genes previously associated with disease pathogenesis were identified in 204 patients (19 %). Patients with variants in *BMPR2* were younger at diagnosis and had more severe pulmonary haemodynamic impairment compared to patients with idiopathic PAH. Novel associations between *BMPR2* variant status and both haemoglobin concentration and white blood cell count were observed. Four patients carried variants in both *SMAD9* and either *BMPR2* or *EIF2AK4*. The clinical significance of this requires further study. Unexpectedly, biallelic variants in *EIF2AK4* were identified in patients with a clinical diagnosis of idiopathic PAH. These patients had a significantly worse prognosis compared to patients with idiopathic PAH. The spectrum of phenotypic, radiological and histological changes associated with biallelic *EIF2AK4* variants was broader than previously recognised.

As these datasets mature, further analyses assessing the response to specific treatments and the outcomes of specific subgroups will be possible.

Preface

I am grateful for the National Institute of Health Research (NIHR) Rare Diseases Translational Research Collaboration for their funding and training over the duration of my PhD. During this time, I have immensely enjoyed working with an incredibly talented group in the Morrell Lab and learned much from them.

This Thesis would not have been possible were it not for the excellent collaboration between many institutions, hospitals and individuals. The parts of the study that I was directly involved with are specified in the Methods section. The genetic and phenotypic data used in the Thesis was generated and collected as part of the NIHR BioResource – Rare Diseases Study (BRIDGE Study) and the Medical Research Council / British Heart Foundation National Cohort Study of Idiopathic and Heritable Pulmonary Arterial Hypertension. I would like to thank everyone involved in both studies, which required the expertise of many individuals to fulfil their objectives. Carmen Treacy, Jennifer Martin and Katherine Yates were instrumental in the organisation and administration of these studies; not to mention their wonderful baking skills!

A special thank you must go to the bioinformatics team, Stefan Gräf, Marta Bleda and Matthias Haimel. They introduced me to a whole new world and patiently taught me the skills needed to complete this PhD. They processed and analysed the whole genome sequencing data presented in this Thesis, as well as creating and maintaining the infrastructure used to collate and process the phenotype data.

I enjoyed collaborating with Nicholas Screaton and Andrew Swift, who reported the CT scans collected as part of the study. It has been an absolute pleasure to work with Mark Southwood, Stephen Preston and Peter Dorfmüller who were involved in preparing and analysing the histological samples.

The engagement of UK pulmonary hypertension centres, their clinicians, nurses and mostly importantly the patients was vital for the study. I would specifically like to thank Joanna Pepke-Zaba and Mark Toshner for their support and the opportunities they have given me at

the Royal Papworth Hospital over many years. My supervisor, Nicholas Morrell, has been hugely influential. He has always been available with helpful advice and suggestions. I would like to thank him for the opportunities he has provided, his guidance during my PhD and for encouraging me to pursue a future academic and clinical career.

Throughout my life my parents have sacrificed much for me. Nothing can repay this unconditional love and support. I hope they will be proud of what I have been able to accomplish over the last few years.

A heartfelt thank you must go to my wife. The birth of our beautiful baby daughter has brought many changes to our lives. Yet, through all this she has remained steadfast with her love, support and understanding, without which I would not have succeeded.

Contents

Tables

Figures

Abbreviations

Activating transcription factor 4	ATF4
Activin A receptor like type 1	ACVRL1
Alanine transferase	ALT
Alkaline phosphatase	ALP
Alveolar volume	VA
Anaerobic threshold	AT
Anti-extractable nuclear antigen antibody	anti-ENA
Anti-neutrophil cytoplasmic antibody	ANCA
Anti-nuclear antibody	ANA
Antibody	Ab
Apolipoprotein E	ApoE
Aquaporin 1	AQP1
Arterial blood gas	ABG
Aspartate transaminase	AST
Base pair	bp
BioResource – Rare Diseases Study	BRIDGE Study
Blood pressure	BP
Body mass index	BMI
Body surface area	BSA
Bone morphogenetic protein	BMP
Bone morphogenetic protein type 2 receptor	BMPR2
Brain natriuretic peptide	BNP
C-reactive protein	CRP
Calcium channel blocker	CCB
Carbon dioxide	CO_2
Carbon dioxide output	VCO_2
Cardiac index	CI
Cardiac output	CO
Cardiopulmonary exercise test	CPET
Caveolin 1	CAV1
CC chemokine ligand 2	CCL2
Cerebellin 2	CBLN2
Chest radiograph	CXR
Chronic obstructive pulmonary disease	COPD
Chronic thromboembolic pulmonary hypertension	CTEPH
Cluster of differentiation	CD
Combined annotation dependent depletion	CADD
Computed tomography	CT

Computer tomography pulmonary angiography	CTPA
Copy number variation	CNV
Cyclic adenosine monophosphate	cAMP
Cyclic guanosine monophosphate	cGMP
Cytoplasmic tail	CT
Diacylgycerol	DAG
Diastolic pulmonary artery pressure	dPAP
Dideoxynucleotides	ddNTPs
Diffusing capacity for carbon monoxide	D_LCO
Diffusing capacity for nitric oxide	D_LNO
Double stranded DNA	dsDNA
Electrocardiogram	ECG
Electronic case report form	eCRF
End tidal partial pressure	PET
Endoglin	ENG
Endothelin receptor A and B	ET-A and ET-B
Endothelin receptor antagonist	ERA
Estimated glomerular filtration rate	eGFR
Ethylenediaminetetraacetic acid	EDTA
Eukaryotic translation initiation factor 2	eIF2
Eukaryotic translation initiation factor 2 alpha kinase 4	EIF2AK4
Eukaryotic translation initiation factor 2 alpha kinase general control nonderepressible 2	GCN2
Eukaryotic translation initiation factor 2 alpha subunit	eIF2α
Exome Aggregation Consortium	ExAC
Extracellular domain	ECD
Forced expiratory volume in 1 second	FEV_1
Forced vital capacity	FVC
Free thyroxine	FT4
Genome Reference Consortium	GRC
Genome wide association study	GWAS
Growth differentiation factor 2	GDF2
Guanine – cytosine	GC
Guanine nucleotide exchange factor (eukaryotic translation initiation factor 2B)	eIF2B
Guanosine-5'-diphosphate	GDP
Guanosine-5'-triphosphate	GTP
Haematocrit	HCT
Haemoglobin	Hb
Heart failure with preserved ejection fraction	HFpEF
Heart rate	HR

Hereditary haemorrhagic telangiectasia	HHT
High density lipoprotein	HDL
High resolution computed tomography	HRCT
High sensitivity c-reactive protein	hsCRP
Human immunodeficiency virus	HIV
Hypoxia-inducible factor 1α	HIF1α
Identifiers	IDs
Inositol triphosphate	IP3
Interleukin	IL
International Classification of Diseases	ICD
Interquartile range	IQR
Interstitial lung disease	ILD
Jugular venous pressure	JVP
Krüppel-like factor 2	KLF2
Left atrium	LA
Left ventricle	LV
Left ventricular ejection fraction	LVEF
Left ventricular end diastolic pressure	LVEDP
Left ventricular end diastolic volume index	LVEDVI
Logarithm of the odds	LOD
Magnetic resonance imaging	MRI
Mean pulmonary artery pressure	mPAP
Micro ribonucleic acid	miRNA / miR
Minor allele frequency	MAF
Mixed venous oxygen saturation	S_vO_2
Monocyte chemoattractant protein 1	MCP1
Mothers against decapentaplegic homolog	SMAD
Multiplex ligation-dependent probe amplification	MLPA
N terminal pro brain natriuretic peptide	NT-ProBNP
National Cohort Study of Idiopathic and Heritable Pulmonary Arterial Hypertension	Cohort Study
National Institute of Health (US)	NIH
National Institute of Health Research	NIHR
Next generation sequencing	NGS
Nitric oxide	NO
Non-coding	NC
Overnight sleep study	OSS
Oxygen	O_2
Oxygen desaturation index	ODI
Oxygen uptake	VO_2
Oxygen utilisation efficiency slope	OUES

Partial pressure of oxygen in arterial blood	P_aO_2
Parts per million	ppm
Peak oxygen consumption	VO_2 max
Peripheral arterial oxygen saturation	S_aO_2
Peroxisome proliferator-activated receptor-γ	PPARγ
Phosphodiesterase 5 inhibitors	PDE5i
Polymerase chain reaction	PCR
Polymorphism Phenotyping v2	PolyPhen-2
Potassium Two Pore Domain Channel Subfamily K Member 3	KCNK3
Primary pulmonary hypertension 1	PPH1
Principal component analysis	PCA
Protein kinase A	PKA
Protein truncating variant	PTV
Pulmonary arterial hypertension	PAH
Pulmonary capillary haemangiomatosis	PCH
Pulmonary capillary wedge pressure	PCWP
Pulmonary embolus	PE
Pulmonary hypertension	PH
Pulmonary vascular resistance	PVR
Pulmonary veno-occlusive disease	PVOD
Pyruvate dehydrogenase	PDH
Quality control	QC
Red blood cell distribution width	RDW
Respiratory exchange ratio	RER
Right atrial pressure	RAP
Right atrium	RA
Right bundle branch block	RBBB
Right heart catheterisation	RHC
Right ventricle	RV
Right ventricular ejection fraction	RVEF
Selective serotonin reuptake inhibitor	SSRI
Sequence Alignment/Map	SAM
Sequencing by synthesis	SBS
Serine / threonine kinase domain	PK
Serotonin	5HT
Serotonin transporter protein	SERT
Single nucleotide polymorphisms	SNPs
Single nucleotide variants	SNVs
Six-minute walk test	6mwt
Sorting Intolerant From Tolerant	SIFT
Standard error	SE

Standard operating procedure	SOP
Systemic lupus erythematosus	SLE
Systolic pulmonary artery pressure	sPAP
T-Box 4	TBX4
Thyroid stimulating hormone	TSH
Thyroxine	T4
Topoisomerase (DNA) II binding protein 1	TOPBP1
Total lung capacity	TLC
Transfer coefficient for carbon monoxide	K_{CO}
Transfer RNA	tRNA
Transforming growth factor β	TGFβ
Transmembrane domain	TM
Tricuspid annular plane systolic excursion	TAPSE
Tumour necrosis factor α	TNFα
Ultrasound	US
Variant call format	VCF
Variant Effect Predictor	VEP
Ventilatory equivalents for carbon dioxide	VE/VCO_2
Vital capacity	VC
White blood cells	WBC
Whole exome sequencing	WES
Whole genome sequencing	WGS
Work rate	WR

Introduction

Pulmonary hypertension

Pulmonary hypertension (PH) is a disorder of the pulmonary circulation defined by a mean pulmonary artery pressure (mPAP) ≥ 25 mmHg measured when supine and at rest by right heart catheterization (1). The 1st World Symposium on Pulmonary Hypertension classified the disorder into 3 broad groups primary, secondary and associated PH (2). The disorder was reclassified in 1998, during the 2nd World Symposium on Pulmonary Hypertension, into 5 groups that continue to form the basis of the latest clinical classification. These groups were thought to share pathophysiological processes, clinical characteristics and similar treatment options. The 3rd World Symposium on Pulmonary Hypertension, in 2003, further amended the classification. The term "primary pulmonary hypertension" was replaced by idiopathic pulmonary arterial hypertension (PAH) (3). This classification was reviewed in 2014 at the 5th World Symposium on Pulmonary Hypertension and further refinements were made based on recent advances in the understanding of the disease, including novel genetic and environmental causes of the disease (4).

Clinical classification

The current clinical classification of PH categorizes the disorder into five groups (1). Group 1 of the current classification system is designated as PAH. The haemodynamic definition of PAH is a mPAP ≥ 25 mmHg and a pulmonary capillary wedge pressure (PCWP) ≤ 15 mmHg, indicating that the pathology is pre-capillary. Subtypes are defined on the presence or absence of associated conditions, exposures to drugs and toxins as well as the presence of pathogenic variants in disease associated genes. Associated conditions include congenital heart disease, connective tissues disease, human immunodeficiency virus (HIV), portal hypertension and schistosomiasis. Exposures to specific drugs and toxins, such as appetite suppressants (e.g. dexfenfluramine), dasatinib and methamphetamines have also been implicated in disease pathogenesis (5-7).

Table 1. Clinical classification of pulmonary hypertension	
Group	**Name**
1	***PAH***
1.1	Idiopathic PAH (no identifiable cause and no family history)
1.2	Heritable PAH (confirmed family history of PAH and / or a pathogenic variant in a disease associated gene): 1.2.1 *BMPR2* 1.2.2 Variants in other genes
1.3	Drug and toxin induced
1.4	Associated with: 1.4.1 Connective tissue disease 1.4.2 HIV 1.4.3 Portal hypertension 1.4.4 Congenital heart disease 1.4.5 Schistosomiasis
1'	***PVOD / PCH***
1'.1	Idiopathic PVOD / PCH
1'.2	Heritable PVOD / PCH: 1'.2.1 Biallelic *EIF2AK4* variants 1'.2.2 Other variants
1'.3	Drug and toxin induced
1'.4	Associated with: 1'.4.1 Connective tissue disease 1'.4.2 HIV
1''	***Persistent pulmonary hypertension of the newborn***
2	***Pulmonary hypertension due to left heart disease***
3	***Pulmonary hypertension due to lung disease or hypoxia***
4	***CTEPH and other pulmonary vascular obstructions***
5	***Multifactorial or unclear mechanism***
Adapted from Galie et al. 2016 (1). PAH – pulmonary arterial hypertension, *BMPR2* – bone morphogenetic protein receptor type 2, HIV – human immunodeficiency virus, PVOD / PCH	

– pulmonary veno-occlusive disease / pulmonary capillary haemangiomatosis, *EIF2AK4* – eukaryotic translation initiation factor 2 alpha kinase 4, CTEPH – chronic thromboembolic pulmonary hypertension.

Heritable PAH, as defined during the 5th World Symposium on PAH, groups together patients with a family history of the disease and / or a pathogenic variant in a disease associated gene. Whereas, idiopathic PAH is a diagnosis of exclusion and by definition the aetiology is uncertain. It requires careful exclusion of the associated conditions and the absence of a family history of the disease and / or a variant in a disease associated gene. However, studies have shown that approximately 20 % of patients with sporadic idiopathic PAH have variants in the gene encoding the bone morphogenetic protein receptor type 2 (BMPR2) (8-16). In comparison, up to 80 % of patients with a family history of the disease have variants in *BMPR2* (17-19).

Group 1' (1 prime) of the current classification is pulmonary veno-occlusive disease (PVOD) / pulmonary capillary haemangiomatosis (PCH) (1). This is an amendment from the previous classification, where it was part of Group 2. It was reclassified to recognise the histological and pathophysiological similarities between PAH and PVOD / PCH. Similar to PAH, PVOD / PCH is also subdivided into idiopathic disease, heritable disease (caused by autosomal recessive variants in the gene eukaryotic translation initiation factor 2 alpha kinase 4 [*EIF2AK4*], that encodes the protein eukaryotic translation initiation factor 2 alpha kinase general control nonderepressible 2 [GCN2]), and cases associated with connective tissue diseases, HIV, and ingestion of drugs or toxins (20-22).

Idiopathic and heritable pulmonary arterial hypertension

Epidemiology

Regulation (EC) No. 141/2000 of the European Parliament from the Council on orphan medicinal products defines a rare disease as pathological condition with a prevalence of less than 5 in 10,000 (23). While the Orphan Drug Act in the United States defines a rare disease as affecting less than 200,000 patients in the United States (24). By either definition, even together, idiopathic and heritable PAH can be considered a rare disease.

Ling et al., using data collected between 2007 and 2009 from all the UK National Pulmonary Hypertension Centres, estimated the incidence of idiopathic, heritable and anorexigen associated PAH at 1.1 per million per year (25). A more recent, but comparable study from the United States (REVEAL Study French Comparison Cohort [i.e. patients with a PCWP ≤ 15 mmHg]) estimated an incidence of 0.9 per million per year (26). Whereas, in a French cohort from 2002 to 2003, the estimated incidence of Group 1 PAH was 2.4 per million of the population per year (27). The incidence of just idiopathic or heritable PAH was not reported.

In the UK, Ling et al. estimated the prevalence of idiopathic, heritable and anorexigen induced PAH to be 6.6 per million of the population (25). With increasing awareness of the disease and longer survival, the UK estimates have increased. The UK National Pulmonary Hypertension Audit 2015, estimated the prevalence of PAH in the UK was now 16 per million of the population (28). While in the United States, prevalence was 10.6 per million of the population from the REVEAL Study (26). Differences between countries may in part be explained by the nature and access to healthcare systems, in particular, whether or not care is centralised to specialist hospitals as it is in the UK and France.

Most contemporary disease registries have not systematically screened for variants in genes previously associated with PAH in their cohorts. The exception to this is in France, where all patients with idiopathic, heritable and anorexigen associated PAH are offered clinical genetic testing (29). Girerd et al. report that in their national cohort of adult PAH patients who underwent genetic screening 80 % had a diagnosis of idiopathic PAH with no known family history of the disease. In this group 17 % of patients had a pathogenic variant in either *BMPR2*,

ACVRL1, *ENG*, *KCNK3*, *CAV1* or *SMAD9*. The majority carried variants in *BMPR2* (15 % of those with idiopathic PAH). Whereas, in patients with a family history of the disease, 89 % carried a disease associated variant. Evans et al. collated the results of eight studies (1550 patients) systematically evaluating *BMPR2* variants in patients with idiopathic, heritable and anorexigen associated PAH (30). They reported that 17 % of patients with idiopathic PAH carried a *BMPR2* variant. In comparison, 82 % of those with a family history of the disease carried a *BMPR2* variant.

Pathophysiology

Genetics basis

A genetic susceptibility to develop PAH has been recognised since the first case reports of the disease occurring within families (31). In 1927 Clarke et al. described two sisters who died of right ventricular failure and proposed "familial" events as a cause of the disease. Subsequently in 1954, Dresdale et al. first described a family with confirmed elevation of their pulmonary artery pressures and suggested a "familial tendency" for "primary pulmonary hypertension" (32). While, Melmon and Braunwald described a three-generation family in which five members of the family were diagnosed with "primary pulmonary hypertension" (31). From this one family, the authors postulated an autosomal dominant mode of transmission with variable penetrance. They also suggested the possibility of genetic anticipation as a younger age of disease onset was reported in the third generation. Most studies of families with PAH concluded that an autosomal dominant mode of inheritance with reduced penetrance was characteristic (33). However, in one family an autosomal recessive mode of transmission was suggested as the affected cases were restricted to just one generation (34). In this family two of the three patients had digital clubbing and pleural effusions, possibly suggestive of PVOD / PCH. However, no histological assessment of the pulmonary veins or capillaries were provided.

Bone morphogenetic protein receptor type 2

In the late 1990s genetic analysis of families with PAH and an apparent autosomal dominant pattern of inheritance suggested that the gene responsible for the disease lay around 2q33, the locus was termed primary pulmonary hypertension 1 (PPH1) (35, 36). Two independent groups assessed the likelihood of linkage (logarithm of the odds score [LOD score]) between

the phenotype and genetic markers across the genome (37). In stark contrast to modern genome wide association studies (GWAS) just 260 and 370 such genetic markers were assessed in the two studies respectively. Follow up studies were then able to use microsatellite polymorphisms to narrow the area in which the responsible gene lay. Sequencing of candidate genes in this area led to the identification of heterozygous protein altering variants in *BMPR2* in the majority (88 % and 47 %) of families studied (18, 19). Following the identification of heterozygous *BMPR2* variants in heritable PAH, pathogenic variants were also reported in 26 % of PAH patients with apparent sporadic disease (i.e. with no confirmed family history) (8).

Following these early studies many other studies have reported protein truncating variants (PTVs), missense variants and splice site variants in *BMPR2* in cohorts of patients with PAH (17, 38). Additionally, multiplex ligation-dependent probe amplification (MLPA) (and prior to this Southern blotting), identified copy number variation (CNVs; deletions or duplications) in *BMPR2* that could not be identified through traditional Sanger sequencing alone (39-41). A recent review by Machado et al. collated 384 distinct *BMPR2* variants, as well as variants in the other 9 genes currently associated with PAH (42).

More recently, a variant in the *BMPR2* promoter region (c.-699G>A), altering transcription factor binding and reducing *BMPR2* transcription has been reported (43, 44). Viales et al. suggested that the promoter variant co-occurred with a *BMPR2* deletion in a patient with PAH possibly influencing penetrance. Other candidate promoter variants have also been identified, that *in-vitro* appear to reduce *BMPR2* transcription, but further work is required to assess the impact and significance of these variants (45). A cryptic translation start site (c.-944_5GC>AT) has also been described in a family with PAH but no protein coding sequence variants in *BMPR2* (46).

Nearly all protein coding sequence variants in *BMPR2* have been found in a heterozygous state, suggesting haploinsufficiency as the mechanism of action (17). One patient, with early onset disease, has been reported to have two missense variants on separate alleles (47). However, it is generally accepted that loss of both *BMPR2* alleles is fatal in utero.

Observing 53 families with heritable PAH and *BMPR2* variants, Larkin et al., were able to show that genetic anticipation was not a feature of heritable PAH caused by *BMPR2* variants (48). They reported that anticipation was an artefact of not following up unaffected family members for sufficient time for the disease to manifest. Furthermore, by assessing the frequency of heritable PAH amongst siblings with *BMPR2* variants they calculated the penetrance of *BMPR2* variants as 27 %. However, there were significant gender differences with penetrance amongst females reported at 42 % and penetrance amongst males as 14 %. This reduced penetrance suggests that a second genetic or environmental "hit" is required to develop disease.

Several mechanisms have been proposed to explain the reduced penetrance. Hamid et al. demonstrated that patients with heritable PAH and *BMPR2* PTVs have reduced levels of their wildtype *BMPR2* transcript compared to unaffected relatives with the same PTV (49). However, this was only tested in patient derived lymphoblastoid cell lines. Transcription of *BMPR2* has been shown to be regulated by several different processes, including inflammation, serotonin signalling and sex hormones; these are discussed in more detail below.

Alternative splicing of *BMPR2*, resulting in the loss of exon 12, was shown to be more common in heritable PAH patients with *BMPR2* variants compared to unaffected *BMPR2* variant carriers (38). This study was also carried out in lymphoblastoid cell lines, which may not recapitulate endothelial cell biology fully (https://gtexportal.org/home/gene/BMPR2). These studies are all in keeping with the concept that additional hits to *BMPR2* signalling are required to manifest disease. This is supported by reduced *BMPR2* expression in endothelial cells from plexiform and concentric lesions from patients with idiopathic PAH, PAH associated with congenital heart disease and scleroderma, as well as patients with chronic thromboembolic pulmonary hypertension (CTEPH) who carry no identifiable *BMPR2* variants (50, 51).

Another mechanism proposed to explain the reduced penetrance of *BMPR2* variants is the occurrence of somatic mutations to cause the loss of the wildtype *BMPR2* allele. However, analysis of microsatellite instability in plexiform lesions dissected from 7 patients with

heritable PAH did not provide any evidence for this (47). Although, Aldred et al. did demonstrate in pulmonary artery endothelial cells in a patient with a germline mutation of *BMPR2*, there was a somatic deletion of chromosome 13 resulting in the loss of mothers against decapentaplegic homolog (*SMAD*) 9, which is part of the *BMPR2* signalling pathway (52).

The *BMPR2* gene consists of 191,442 base pairs (bp) and is divided into 13 exons that are all transcribed in the canonical transcript. It encodes the 1038 amino acid type 2 transforming growth factor β (TGFβ) receptor, bone morphogenetic protein receptor type 2 (53). Five distinct regions / domains are recognised: a signal sequence (exon 1), the extracellular ligand binding domain (exons 2 and 3), a transmembrane domain (exons 4 and 5), a serine / threonine kinase (exons 6 to 11) and a large cytoplasmic tail that may interact with the cytoskeleton and cSrc (exons 12 and 13) (54-56). Disease associated variants have been reported in all exons (42). Variants in the cytoplasmic tail may not disrupt canonical *BMPR2* signalling but may affect interactions with the cytoskeleton and cSrc (55, 57-59).

BMPR2 forms a heterodimeric complex with a TGFβ type 1 receptor, and this determines ligand specificity. In endothelial cells, BMPR2 forms a complex with activin A receptor like type 1 (ACVRL1) and binds specifically to bone morphogenetic protein 9 (BMP9; encoded by the gene, growth differentiation factor 2 [*GDF2*]), and bone morphogenetic protein 10 (BMP10) (60). Endoglin (ENG) acts as an accessory receptor promoting ACVRL1 signalling over competing ALK5 signalling (61). Ligand binding to BMPR2 results in phosphorylation and activation of ACVRL1 leading to downstream signalling through receptor-regulated SMAD proteins. In particular, ACVRL1 phosphorylates SMAD 1, 5 and 8 (encoded by *SMAD1*, *SMAD5* and *SMAD9* respectively). These activated receptor-regulated SMAD proteins form heteromeric complexes with SMAD4 and are translocated to the nucleus where they can bind to SMAD binding elements to promote transcription. Variants in genes other than *BMPR2* involved in this signalling cascade have also been implicated in heritable PAH disease pathogenesis.

Activin A receptor like type 1

ACVRL1 encodes a type 1 TGFβ receptor that is expressed on endothelial cells (62). In complex with BMPR2, ACVRL1 activation leads to phosphorylation of the receptor-regulated SMADs (SMADs 1, 5, 8). Variants in *ACVRL1* were initially reported in hereditary haemorrhagic telangiectasia type 2 (HHT2 and also known as Rendu-Osler-Weber syndrome), a highly penetrant autosomal dominant disease associated with the formation of mucocutaneous telangiectases and arteriovenous malformations (63). Vorselaars et al. recently estimated that approximately 18 % of patients with HHT have evidence of pulmonary hypertension when assessed by echocardiography (64). This may be related to an elevated PCWP, high pulmonary artery flow secondary to arteriovenous malformations or due to PAH (65).

Variants in *ACVRL1* are reported in heritable and idiopathic PAH usually in association with HHT (66, 67). There are also reports of paediatric onset PAH patients with variants in *AVCRL1* who do not have the clinical features of HHT (11, 13). These patients may manifest features of HHT later in life.

Endoglin

Endoglin, encoded by *ENG*, is an accessory receptor expressed on endothelial cells that enhances receptor-regulated SMAD signalling through the BMPR2-ACVRL1 complex. Similar to *ACVRL1*, variants in *ENG* are associated with HHT (HHT type 1) (68). HHT type 1 and HHT type 2 show phenotypic differences. The frequency and locations of arteriovenous malformations differ between the two forms of HHT. Furthermore, in a small study of patients with HHT and confirmed pulmonary hypertension (n = 28), *ENG* variants were less common than *ACVRL1* variants (6 *ACVRL1* variants, 1 *ENG* variant and 1 *BMPR2* variant; the identities of these variants were not stated to allow further critique) (65). Variants in *ENG* have been reported in patients with PAH and HHT (11, 67, 69). Although in a number of these patients, other risk factors for pulmonary hypertension were also present including anorexigen use, congenital heart disease and high pulmonary artery flows.

Other TGFβ pathway genes

Variants in other genes in the TGFβ signalling pathway have also been reported. In some cases, functional data and demonstration of familial co-segregation of the variants have not been provided to confirm a causal role in PAH pathogenesis. Further study is hindered by the fact variants in these genes are extremely rare. Nasim et al. looked for rare variants in SMAD proteins in patients with PAH and no variants in *BMPR2* or *ACVRL1*. They reported 2 variants in *SMAD4* (associated with HHT type 3), 1 variant in *SMAD1* and 1 variant in *SMAD9* from a mixed cohort of 324 patients with PAH (70). The four patients with variants all had a diagnosis of idiopathic PAH. The *SMAD9* variant carrier had paediatric onset disease. Shintani et al. also described a paediatric onset PAH patient with a protein truncating variant in *SMAD9* resulting in loss of the ACVRL1 phosphorylation site; the variant was not present in their unaffected father (71).

Chida et al. described two paediatric patients with idiopathic PAH carrying variants in *BMPR1B*, an alternative TGFβ type 1 receptor (72). Although one variant, p.F392L, appeared to promote SMAD8 phosphorylation in the presence of BMP4 and it was also found in the unaffected father of the proband, casting doubt on its role in disease pathogenesis.

More recently, a case report of a homozygous missense variant in *GDF2* causing childhood onset PAH with no features of HHT has been published (73). Both unaffected parents were heterozygous for the variant suggesting an autosomal recessive pattern of inheritance. *GDF2* variants, in a heterozygous state, have been reported in three patients with capillary malformations and a phenotype similar to HHT, although only two variants appeared to impact the signalling pathway (74). No information regarding the presence or absence of pulmonary hypertension in these two patients were provided.

Potassium two pore domain channel subfamily K member 3

Potassium two pore domain channel subfamily K member 3 (*KCNK3*) encodes a potassium channel. It was initially identified by Ma et al. through whole exome sequencing (WES) of a single family in which five members had heritable PAH and no identifiable variants in genes previously associated with the disease (75). An autosomal dominant pattern of inheritance

was described in the family. By identifying variants that were 1) rare, 2) found in affected family members, 3) found in a heterozygous state (given the autosomal dominant mode of transmission) and 4) predicted to be deleterious, they were able to narrow the search to 24 variants. The variant in *KCNK3* was considered further as the gene had previously been associated with maintaining pulmonary artery vascular tone (76). Screening of patients with no variants in genes previously associated with disease pathogenesis revealed *KCNK3* variants in 1 % of patients with idiopathic PAH and 3 % of probands with heritable PAH (i.e. a family history of the disease). The patients were relatively young at disease onset (aged 8 to 44 years old) and did not respond to acute pulmonary vasodilator challenges. Incomplete penetrance was suggested by the presence of variant carriers who had not developed the disease by their 40th decade. *KCNK3* variants have subsequently been reported in other PAH cohorts (77, 78). Interestingly, Tejedor et al. include a Romani family in which the index case carried a *KCNK3* variant in a homozygous state and was diagnosed with PAH at just 2 months of age. Both parents were found to be heterozygotes and the mother developed PAH a year after giving birth to the index case (77).

Caveolin 1

Caveolin 1 (encoded by *CAV1*) is a transmembrane protein integral to the formation of caveolae. Caveolae are flask like invaginations in the plasma membrane that are thought to play a role in promoting receptor mediated signalling and mechanotransduction (79). By employing WES to study a family with heritable PAH, Austin et al. were able to identify a variant in *CAV1* (80). Subsequently, screening for *CAV1* variants in 260 patients with either heritable or idiopathic PAH revealed just one other patient with a de novo frameshift variant and a diagnosis of idiopathic PAH. Thus, variants in *CAV1* as a cause of PAH are extremely rare. Several studies provide descriptions of the same young girl with both idiopathic PAH and congenital generalised lipodystrophy (81-83). Congenital generalised lipodystrophy is associated with absence of adipose tissue, abnormal lipid metabolism and insulin resistance, and has been associated with *CAV1* variants in the absence of pulmonary hypertension (84, 85). Gomez et al. reported another patient with a missense variant in *CAV1,* but this is actually a variant of unknown significance with discordant Sorting Intolerant From Tolerant (SIFT) and Polymorphism Phenotyping v2 (PolyPhen-2) deleteriousness scores (86).

T-Box 4

Microdeletions involving 17q22q23.2 have been reported to cause early onset pulmonary hypertension as well as other congenital abnormalities (87, 88). These microdeletions were postulated to cause haploinsufficiency of *TBX4,* encoding the transcription factor, T-Box 4. Kerstjens-Frederikse et al. found evidence of microdeletions involving 17q22q23.2 in 3 out of 6 paediatric onset cases of idiopathic PAH who also had unexplained mental retardation or dysmorphic features (89). These deletions were not present in the unaffected parents and the area of overlap between all 3 cases contained 6 genes including *TBX4*. Amongst a further 14 paediatric onset cases of idiopathic PAH another 3 patients were found to have deleterious variants in *TBX4*. Reassessment of these patients revealed features consistent with small patella syndrome (hypoplastic or absent patella, a large gap between 1st and 2nd toes and long 3rd and 4th toes), a syndrome previously associated with *TBX4* variants (90). Amongst a cohort of 49 adult patients one patient with idiopathic PAH was found to carry a deleterious *TBX4* missense variant but did not have features of small patella syndrome. The authors found no evidence of PAH amongst 23 patients diagnosed with small patella syndrome. *TBX4* variants have since been described in further cohorts of paediatric and adult onset PAH (91, 92).

Other putative genes

Rare and predicted deleterious variants in several other genes have also been implicated in PAH pathogenesis, however, additional numbers, further phenotypic information and/or functional studies are required to confirm their significance. Biallelic *EIF2AK4* variants were first reported in patients with PVOD / PCH (20, 21). More recently two reports have described biallelic *EIF2AK4* variants in a kindred of Iberian Romani and in two sisters with apparent heritable PAH (22, 93). In the former report, half the patients had a clinical suspicion of PVOD / PCH although no histological confirmation of the diagnoses were presented. In the latter study, both sisters had a low transfer coefficient for carbon monoxide (KCO) but neither had assessment of the lung parenchyma by computer tomography (CT), that may have suggested a diagnosis of PVOD / PCH.

A single case report describes a family with heritable PAH and a missense variant in Krüppel-like factor 2 (*KLF2*) (94). *KLF2* encodes a shear stress induced transcription factor, that can increase expression of aquaporin 1 (*AQP1*) and thereby increase nitric oxide (NO) mediated

vasorelaxation (95, 96). Variants in *KLF2* have also been implicated in splenic marginal zone lymphoma (97).

KCNA5, encodes the α subunit of the voltage gated potassium channel, Kv1.5. Decreased expression of Kv1.5 in pulmonary artery smooth muscle cells from patients with PAH has been observed and results in increased cellular proliferation (98). Remillard et al. identified both common and rare variants in both the protein coding sequence and putative promoter sequence of *KCNA5*, that were predicted to impact channel function or interfere with transcription factor binding (99). They reported that the frequency of two single nucleotide polymorphisms (SNPs) were significantly different between patients who had a vasoreactive response to an acute pulmonary vasodilator challenge and those that did not. While the frequency of another SNP was different between those with a history of exposure to fenfluramine / phentermine and those without. In a separate Spanish cohort, predicted deleterious and rare missense variants, as well as novel splice site variants in *KCNA5* have been reported (100). Furthermore, a case report describes early onset heritable PAH in a patient with both a missense *BMPR2* variant (previously reported in several patients with heritable PAH) and a frameshift variant in *KCNA5* (101). The authors hypothesised that the early age of onset was due to the combination of the two variants. The patient's unaffected mother also carried the *KCNA5* variant.

As well as rare and predicted deleterious variants that are likely causal for PAH, common SNPs have also been implicated in disease pathogenesis usually as disease modifiers. Candidate gene association studies and a single GWAS have assessed the role of common variation in PAH pathogenesis. SNPs, identified through candidate gene approaches, in genes linked to inflammation, metabolism and sex hormone signalling are discussed in more detail in the relevant sections below.

GWAS require large well-defined cohorts, appropriate controls and careful selection of SNPs to take forward into the association study. In the only GWAS in PAH to date, Germain et al. assessed 340 patients with idiopathic or heritable PAH and no identifiable *BMPR2* variants (102). They were unable to identify any SNPs that were significantly different between cases and controls at a genome wide significance level in their discovery cohort. However, they took

forward 319 SNPs that had uncorrected p values less than 6.87 x10^{-4} for further analysis. Two of these SNPs were then also seen in a smaller validation cohort of 285 patients. Both SNPs were downstream of the cerebellin 2 precursor gene (*CBLN2*). *CBLN2* encodes cerebellin-2, which had previously been thought to be expressed by neuronal tissue. Germain et al. showed that cerebellin-2 was expressed at higher levels in pulmonary artery endothelial cells from PAH patients compared to those from controls, but their functional role in PAH pathogenesis remains to be elucidated.

In a small WES study of just 12 patients with idiopathic PAH, common variation in topoisomerase (DNA) II binding protein 1 (*TOPBP1*) was suggested to be associated with disease pathogenesis (103). However, the methodology used in the study to identify the variants was biased and the study itself remains underpowered.

Summary of the genetic basis of PAH

The *BMPR2* signalling pathway is now recognised as a playing a crucial role in disease pathogenesis. *BMPR2* mutations are the commonest cause of heritable PAH. Although BMPR2 is expressed in several cell types, in endothelial cells BMPR2 forms a unique complex with ACVRL1, and the accessory receptor ENG. When bound to the ligands BMP9 and BMP10 this receptor complex initiates a specific signalling cascade of importance to endothelial cell biology. Mutations in genes encoding all these proteins, with the exception of BMP10, have already been identified. This also demonstrates the critical role of the pulmonary endothelial cell in disease pathogenesis.

Additionally, there is an increasing body of evidence to suggest that variation in genes not directly part of the canonical *BMPR2* signalling pathway, and non-genetic factors (which are further discussed below) can modulate this key pathway. This growing understanding of disease pathogenesis has led to the development of interventions which promise to deliver new classes of therapeutics to combat and potentially cure this disease. Figure 1 summarises the *BMPR2* signalling pathway and includes other genetic and non-genetic factors that may influence it.

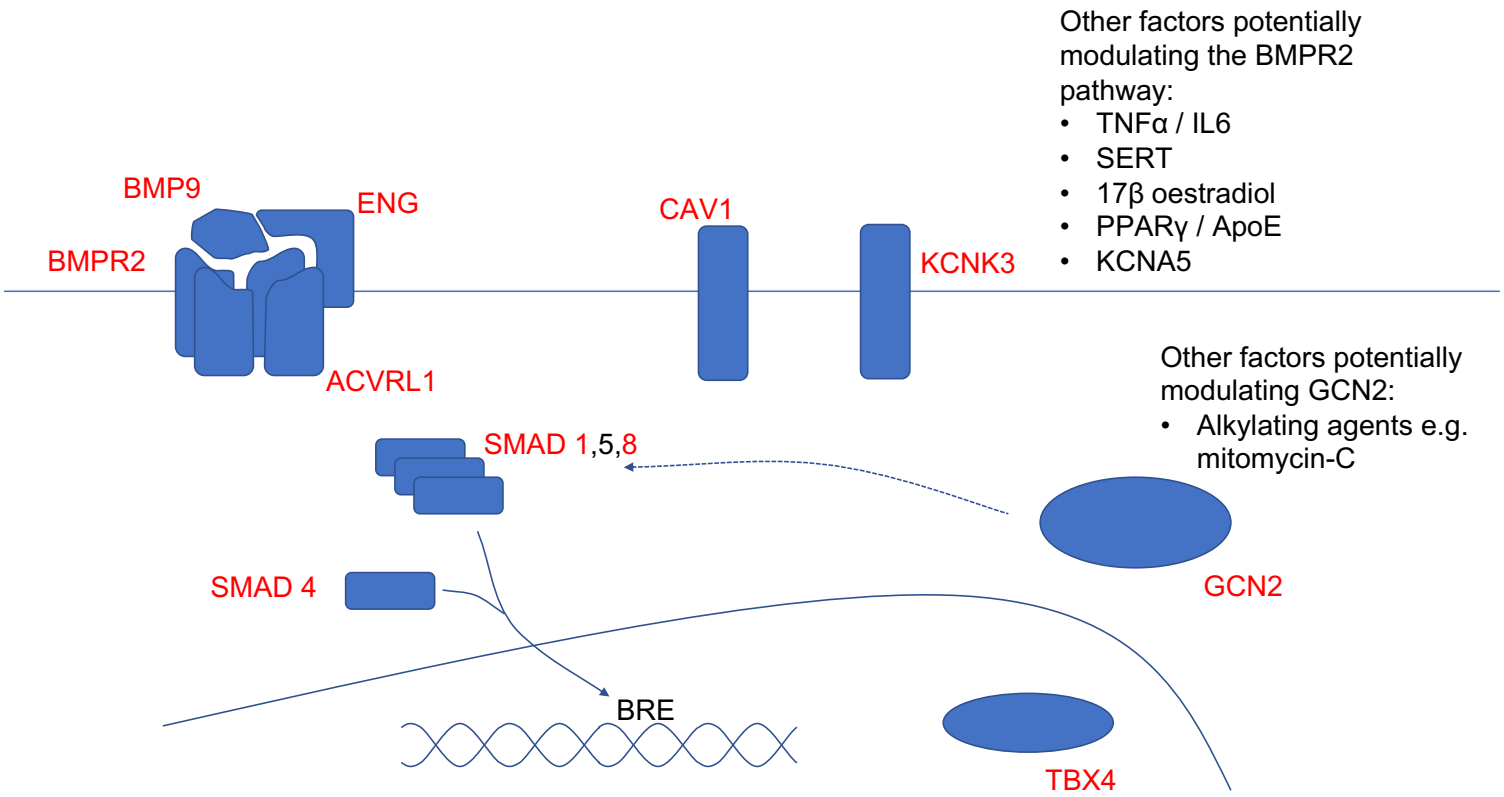


Figure 1. Diagram of the *BMPR2* signalling pathway and other genes previously associated with PAH and PVOD/PCH

The *BMPR2* signalling pathway plays a central role in the pathogenesis of PAH. BMPR2 is a type 2 TGFβ receptor which forms a receptor complex with ACVRL1 on the endothelial cell surface membrane. This receptor complex, along with the co-receptor ENG, have an affinity to the ligands BMP9 and BMP10. Ligand binding leads to phosphorylation of SMADs 1, 5 and 8, which then bind to SMAD 4 before being translocated to the nucleus where they can bind to BMP responsive elements (BRE) to influence transcription. Mutations in many genes contributing to this signalling cascade have already been implicated in the development of PAH (shown in red). Other genes have also been implicated in disease pathogenesis (also labelled in red). CAV1 helps the formation of caveolae on the cell surface membrane where BMPR2 receptors colocalise. KCNK3 is a potassium channel implicated in controlling vascular tone. TBX4 is a transcription factor implicated in paediatric onset PAH. GCN2, encoded by *EIF2AK4*,

has been associated with both PAH and PVOD / PCH. It is part of the integrated stress response. When activated it phosphorylates eIF2α, which leads to an increase in ATF4 transcription and a global reduction in protein translation.

BMPR2 – bone morphogenetic protein type 2 receptor, ACVRL1 – activin A receptor like type 1, ENG – endoglin, BMP9 – bone morphogenetic protein 9, SMAD – mothers against decapentaplegic homolog, BRE – BMP responsive elements, CAV1 – caveolin 1, KCNK3 – potassium two pore domain channel subfamily K member 3, TBX4 - T-Box 4, GCN2 - eukaryotic translation initiation factor 2 α kinase general control nonderepressible 2, TNFα – tumour necrosis factor α, IL – interleukin, SERT – serotonin transporter protein, PPARγ – peroxisome proliferator-activated receptor γ, ApoE – apolipoprotein E, KCNA5 - Kv1.5 α subunit.

Histological descriptions

Remodelling of the pulmonary vasculature was first recognised by Romberg in 1891, when he described sclerosis of the pulmonary arteries. Dresdale et al. associated these changes with "primary pulmonary hypertension". In their first descriptions of "primary pulmonary hypertension" they described widespread but variable "sclerotic changes" throughout the pulmonary circulation (104). They also reported plaque formation in larger vessels, intimal and medial thickening, as well as subintimal proliferation narrowing the vessel lumen of microscopic vessels. Thrombotic lesions at different stages of organisation in both pre and post-capillary microscopic pulmonary vessels were also observed.

Wagenvoort and Wagenvoort described the histological features of lung tissue from 156 patients with "primary pulmonary hypertension" and considered it possible to categorise them into groups: chronic thromboembolism, chronic pulmonary venous hypertension, pulmonary veno-occlusive disease, sarcoidosis, chronic bronchiolitis and emphysema, pulmonary schistosomiasis and "vasoconstrictive primary pulmonary hypertension" (105). "Vasoconstrictive primary pulmonary hypertension" was characterised by intimal fibrosis, plexiform lesions, fibrinoid necrosis and pulmonary arteritis.

Intimal thickening, medial hypertrophy, perivascular inflammation, and the presence of plexiform lesions are now considered key histological features of PAH. Medial hypertrophy is caused by proliferation and hypertrophy of the vascular smooth muscle cells and can correlate with pulmonary vascular resistance (PVR) (106).

Plexiform lesions were once considered pathognomonic of idiopathic and heritable PAH but have also been reported in other forms of PH including CTEPH and PAH associated with congenital heart disease (107-110). Additionally, plexiform lesions are only variably reported in patients with idiopathic PAH, ranging from 20 % to 90 % (106, 109). Plexiform lesions are characterised by endothelial cell proliferation in pulmonary arteries < 100 μm in diameter and can result in near obliteration of the vessel lumen. They were thought to be associated with severe pulmonary hypertension (111).

A more recent histological study, by Ghinga et al., reported pulmonary venous remodelling, in particular smooth muscle hyperplasia in septal veins, in patients with *BMPR2* variants (112). This study, compared 23 patients with *BMPR2* variants to 21 patients with idiopathic PAH and no identifiable disease associated variants. The amount of arterial remodelling, the density of plexiform lesions and the degree of perivascular inflammation was similar between the two groups. However, *BMPR2* variant carriers had increased bronchial artery hypertrophy and dilation compared to patients with idiopathic PAH. Changes to the bronchial circulation were associated with a history of haemoptysis.

Inflammation

Many studies have suggested that inflammation is an important aspect of PAH pathogenesis and an interaction between inflammation and genetic factors has been postulated. Pathological studies have described an abundance of inflammatory cells in the perivascular tissue, particularly around plexiform lesions (106, 113). T cells, B cells and macrophages have been observed. The degree of inflammation within the perivascular space has been associated with the severity of pulmonary hypertension (106). However, no difference in the degree of perivascular inflammation was seen between patients with *BMPR2* variants and patients with idiopathic PAH (112).

Immune dysregulation is thought to be an important process in some forms of PAH such as those associated with HIV infection, systemic sclerosis and systemic lupus erythematosus (SLE). Tamosiuniene et al. reported the development of severe pulmonary hypertension in athymic mice exposed to the vascular endothelial growth factor receptor 2 antagonist, SU5416 (Sugen), alone (unlike in regular mice where hypoxia is also required) (114). They found that immune reconstitution with regulatory T cells (cluster of differentiation [CD] $4^{+}CD25^{hi}$ or $CD4^{+}CD25^{-}$) prior to exposure of Sugen prevented the development of pulmonary hypertension. This was associated with reduced perivascular inflammation (B cells and macrophages), reduced circulating levels of tumour necrosis factor α (TNFα) and interleukin (IL) 6, as well as increased *BMPR2* expression in pulmonary vascular cells.

Serological studies have identified increases in circulating cytokines, such as IL1α, IL1β, IL2, IL4, IL6, IL8, IL10, IL12p70, IL13, TNFα, monocyte chemoattractant protein 1 (MCP1) and CC chemokine ligand 2 (CCL2), in patients with idiopathic or heritable PAH (115-119). No differences in these cytokines have been reported between patients with idiopathic PAH and those carrying *BMPR2* variants (115). Although, preclinical studies have suggested increased IL6 in mice with dominant negative variants in *BMPR2* in smooth muscle cells (120). Some of these cytokines are reported to be prognostic markers in univariate analyses, but not always in multivariate Cox proportional hazards models (115, 119, 121).

Chaout et al. further investigated the link between *BMPR2* and IL6. In a study of patients with chronic obstructive pulmonary disease (COPD), patients with the G / G genotype at position c.-174 in the gene encoding IL6 (G being the major allele) was associated with an increased susceptibility to developing pulmonary hypertension and increased levels of circulating IL6 amongst current smokers (122). Brock et al. proposed that IL6 reduced *BMPR2* function by increasing transcription of the micro ribonucleic acid (miRNA / miR) 17-92, which in turn reduces translation of BMPR2 (123). Preclinical models have also suggested that a "second hit", such as exposure to TNFα, lipopolysaccharide or hypoxia, is required in mice with heterozygous *BMPR2* variants, to develop significant pulmonary hypertension (124-127).
Further indirect evidence for a role of inflammation in PAH arises from the co-occurrence of inflammatory disorders in patients with idiopathic PAH, such as autoimmune thyroiditis (128-130). Patients with an initial clinical diagnosis of idiopathic PAH can develop autoantibodies and connective tissue diseases over time.

Metabolic changes in PAH

Proliferation and hypertrophy of pulmonary artery smooth muscle cells, endothelial cells and fibroblasts are features of the vascular remodelling observed in PAH (106, 131). This is associated with changes to cellular metabolism, similar to that seen in malignant processes. In fact, the proliferative changes associated with PAH, such as plexiform lesions, have been reported as being reminiscent of cancers (113).

The exact cause of this hyperproliferative, anti-apoptotic cellular phenotype remains to be fully elucidated, but mitochondrial dysfunction has been proposed as an important factor. Mitochondria from PAH tissue are hyperpolarised. This can contribute to a proliferative phenotype through activation of transcription factors such as hypoxia-inducible factor 1α (HIF1α) and altered intracellular calcium and potassium concentrations (132-134).

Furthermore, suppression of mitochondrial respiration results in a shift from aerobic respiration / glucose oxidation to an increased reliance on glycolysis even in normoxic conditions (the Warburg Effect), similar to that seen in malignant cells (135). Pyruvate dehydrogenase (PDH) is a critical enzyme in this regard, linking glycolysis to glucose oxidation by playing a role in the conversion of pyruvate to acetyl-CoA, which can enter the citric acid cycle. Pyruvate dehydrogenase kinase (PDK) can reduce PDH activity, increasing glycolysis and reducing glucose oxidation. Dichloroacetate, an inhibitor of PDK, can both prevent and reverse the development of pulmonary hypertension in preclinical models (136-138). The inflammatory cytokine, TNFα, can also reduce PDH activity (139). A consequence of the reduction of PDH activity is increased utilisation of fatty acids in the citric acid cycle and diversion of critical amino acids for cellular proliferation (140). Zhao et al. took a global metabolomics approach to investigate changes in metabolism in the lungs of patients with end stage PAH. Compared to control lung tissue, lung tissue from PAH patients showed a significantly different profile based on unsupervised principal component analysis (PCA). They were able to demonstrate significant differences in metabolites associated with the glycolytic, citric acid cycle and fatty acid oxidation pathways.

The development of insulin resistance is another feature of this metabolic impairment seen in patients with PAH (141-143). Several preclinical studies have suggested links between pulmonary hypertension and insulin resistance. Deficiency or knock out of the ligand activated transcription factor, peroxisome proliferator-activated receptor γ (PPARγ), or its downstream effector, apolipoprotein E (apoE) can cause both insulin resistance and pulmonary hypertension in mice (144, 145). Interestingly deficiency of adiponectin (another protein downstream of PPARγ), is associated with obesity and insulin resistance, and results in the development of pulmonary hypertension in an age dependent manner in mice (146). This may be driven through reduced NO production or increased expression of endothelial

adhesion molecules such as E-selectin. PPARγ is reported to be reduced in the lungs of patients with PAH (147). The importance of insulin resistance in PAH is further demonstrated by the prognostic significance of haemoglobin (Hb) A1C, a marker of insulin resistance and hyperglycaemia (148).

BMPR2 signalling is thought to play a role in cellular metabolism and survival. Early in the disease endothelial cells are considered to have a pro-apoptotic phenotype that results in the selection and expansion of endothelial cells with the anti-apoptotic phenotype (149). *BMPR2* signalling is thought to promote endothelial cell survival, therefore pathogenic variants in *BMPR2* can promote endothelial apoptosis. This potentially acts as the initiating event leading to the development of PAH (150-152). Furthermore, the anti-proliferative effects of *BMPR2* signalling is dependent on PPARγ and ApoE (145). *BMPR2* endothelial cell knockout mice have also been shown to have hyperpolarised pulmonary endothelial mitochondrial membrane potentials and this has been attributed to increased p53 and PPARγ co-activator 1α expression (153).

Comparison of human pulmonary microvascular endothelial cells with and without *BMPR2* variants showed changes across many metabolic pathways. This included increases in metabolites found in the glycolytic pathway and decreases in metabolites associated with fatty acid metabolism and the citric acid cycle (154). Similar changes have been shown in the right ventricle of patients with *BMPR2* variants and this is thought to cause impaired right ventricular function through increased cardiac lipid deposition (155).

Serotonin

Several lines of evidence suggest serotonin (5HT) is involved in disease pathogenesis, particularly mediating proliferation of vascular smooth muscle and vasoconstriction (156, 157). The relationship between 5HT and PAH was first suggested following an epidemic of PAH associated with anorexigen use in the late 1960s (5, 158). The latest guidelines now recognise aminorex, fenfluramine, dexfenfluramine, benfluorex and selective serotonin reuptake inhibitors (SSRIs) as “definite” causes of PAH (1, 5, 6, 159-162). All of these modulate the 5HT pathway. The histological appearances of the pulmonary vasculature in patients

exposed to aminorex are similar to that seen in other forms of PAH and include the presence of plexiform lesions. However, there are some notable phenotypic features of patients exposed to anorexigens. There is evidence of disease reversibility and good long-term outcomes in patients with anorexigen associated PAH, unlike patients with other forms of PAH (163). Administration of aminorex to rats failed to cause the disease and only 2 % of patients taking the drug developed PAH, suggesting a genetic predisposition, particularly in women, is required (163, 164).

The serotonin transporter protein (SERT) is required for the proliferative effects of 5HT on pulmonary vascular smooth muscle (165-167). Anorexigens can inhibit SERT and promote the release of indolamine, both of which increase extracellular 5HT (161, 168). Furthermore, anorexigens can be a substrate for SERT, leading to intracellular 5HT like actions (169). In a meta-analysis, an insertion / deletion polymorphism (L / L genotype) in the promoter region of *SERT* was associated with increased *SERT* expression and PAH (170-172). In one of the original studies included in the meta-analysis, patients with *BMPR2* variants and the L / L genotype had a younger age at diagnosis (171).

Methamphetamines are also associated with the development of PAH and can interact with SERT at high doses (168). Although methamphetamines and amphetamines can also exert their effects through other mechanisms such as by increasing oxidative phosphorylation leading to increased reactive oxygen species and deoxyribonucleic acid (DNA) damage (173).

Further evidence for the role of 5HT in PAH comes from elevated plasma 5HT in PAH patients, although this is not a consistent finding (174, 175). This was thought to arise due to abnormal platelet handling. Additionally, SSRIs have been associated with persistent pulmonary hypertension of the newborn (176).

Oestrogen

Interest in the role of the sex hormones in PAH pathogenesis arose from the female sex bias seen in many forms of PAH (5, 177). Oestrogen and its metabolites have been shown to influence pulmonary artery endothelial and smooth muscle cells, potentially explaining the increased risk of disease in females. Cytochrome P450 1B1 is involved in the oxidation of

oestrogen to 2- and 4- hydroxyestrogens. While hydroxylation of oestrogen results in the formation of 16α hydroxyestrone, which is more potent at causing cellular proliferation through activation of the oestrogen receptors compared to other oestrogen metabolites. Polymorphisms in cytochrome P450 1B1, resulting in reduced transcription, have been reported in female patients with heritable PAH (178). In keeping with this the ratio of urinary 2-hydroxyestrogen to 16α hydroxyestrone was reduced in a small subgroup.

The *BMPR2* promoter sequence also contains an oestrogen receptor binding site that reduces *BMPR2* transcription (179). In human pulmonary artery smooth muscle cells, *BMPR2* expression was greater in male derived cells compared to female derived cells (180). In this study, 17β oestradiol (the main oestrogen produced prior to menopause) increased cellular proliferation, providing a potential mechanism to explain the female bias in PAH registries and increased penetrance of *BMPR2* variants in women. Furthermore, White et al. showed that in *SERT* overexpressing mice only females developed pulmonary hypertension and this was abolished by oophorectomy (181). They went on to show that administration of 17β oestradiol permitted the development of pulmonary hypertension in the oophorectomized, *SERT* overexpressing mice. This may explain, at least in part, the female bias seen in anorexigen associated PAH.

Paradoxically, despite the female gender bias in PAH, female patients have a better prognosis compared to male patients (182, 183). This may be related to the effect of sex hormones on the right ventricle (184, 185). Several preclinical studies suggest that oestrogen metabolites improve right ventricular remodelling in pulmonary hypertension (181, 186).

In mice models of pulmonary hypertension, treatment with antagonists of the oestrogen receptor α or anastrozole, an aromatase inhibitor, led to increased *BMPR2* expression and prevented or reduced the development of pulmonary hypertension (187, 188). A recent study also showed that the combination of anastrozole and fulvestrant (a selective oestrogen receptor degrader), can prevent and even reverse pulmonary hypertension in a mouse model with an inducible *BMPR2* variant (189). This was associated with increases in PPARγ and improvements in insulin resistance. Furthermore, a recently published small phase 2 study of 18 PAH patients showed that anastrozole could increase six-minute walk test (6mwt) distance

after 3 months of therapy (190). Further work is required to assess the benefit of such an intervention, especially given the potentially protective effect that 17β oestradiol may have on the right ventricle.

Demographics

An increasing awareness of the disease amongst both primary and secondary care physicians in recent years has resulted in increasing referrals, particularly of older patients, to specialist centres (28). Compared to early PAH registries, such as the US National Institute of Health (NIH) Study, analysis of the REVEAL registry demonstrated that the average age at diagnosis was now significantly older (36.4 years vs. 44.9 years) (26, 177). Similar trends have been seen in the UK, where Ling et al. showed that, between 2001 and 2009, the mean age of diagnosis increased amongst patients diagnosed with idiopathic, heritable and anorexigen induced PAH (25). This was associated with an increased body mass index (BMI), more comorbidities and a reduction in the K_{CO}. Recent registries from Brazil and China, show that the mean age of diagnosis for patients diagnosed with idiopathic PAH is still less than 40 years of age (191, 192). Taken together, this may suggest that the pathogenesis of disease in patients diagnosed with apparent idiopathic PAH in modern cohorts from more economically developed countries may be more heterogeneous, with increasing numbers of patients with cardiovascular comorbidities (193). In the prospective European Compera registry the median age of diagnosis for patients with idiopathic PAH was 71 years (194).

A bias in the proportion of female to male patients is seen in all forms of pulmonary arterial hypertension. In the US NIH Study, the ratio of female to male patients diagnosed with idiopathic or heritable PAH was 1.7 to 1. In modern registries, the bias is sometimes even more prominent but has been variable. The REVEAL Study had the highest ratio 3.6 to 1 (3.8 to 1 in the comparable $REVEAL^{NIH}$ cohort). While UK, French and European registries have ratios between 1.6 and 2.3 to 1 (25-27, 194).

The only recent study to assess the ethnicities of patients with PAH was the REVEAL Study (26). Frost et al. report that the frequency of white Caucasians was as expected, whereas patients of Afro-Caribbean ancestry were over-represented and those of Asian or Hispanic ancestry under-represented. The authors comment that this may be due to ascertainment

biases and disparities in access to health care. Additionally, Kawut et al. identified Asian or African-American ethnicity to be a poor prognostic marker (195).

Clinical features and assessments

Symptoms associated with PAH are non-specific and usually insidious in onset. These factors combined with the rarity of the disease and lack of familiarity amongst health care providers, can delay clinical suspicion of the disease, referral to a specialist centre and definitive management. Analysis of the REVEAL Study population revealed that younger patients and those with less severe pulmonary haemodynamics were more likely to have a delay from symptom onset to diagnosis (defined as more than 2 years) (196). Smaller studies have quantified that the mean duration between symptom onset and diagnosis was between 2.3 and 3.9 years even in modern healthcare systems (27, 197, 198). It has been presumed that older patients are more likely to be offered early echocardiography to assess left ventricular or valvular function and are consequently referred to specialist centres earlier.

The most common symptom associated with PAH is unexplained exertional dyspnoea, with 98 % of the UK cohort reported by Ling et al. presenting with dyspnoea (25, 177). Enlargement of the pulmonary artery may result in extrinsic compression of the other structures in the mediastinum leading to cough (compression of main bronchi), hoarse voice (compression of left recurrent laryngeal nerve) and chest pain (compression of left main coronary artery) (199). Haemoptysis is reported particularly in those with *BMPR2* variants (112). Symptoms associated with worsening right ventricular function include palpitations, postural dizziness and syncope on exertion. Ling et al. reported that younger patients were more likely to present with syncope, perhaps consistent with their later presentation to health care providers (25). Syncope is associated with poor prognosis as it is a reflection of impaired cardiac function (200). Other non-specific symptoms include fatigue and abdominal discomfort from ascites.

Clinical signs elucidated from patients with PAH are also non-specific and unreliable for confirming the diagnosis (201). The most frequent signs are related to right ventricular hypertrophy and failure: parasternal heave, pan systolic murmur loudest on inspiration, loud pulmonary component of the second heart sound, elevated jugular venous pulse with a large

v wave, ascites and peripheral oedema. Digital clubbing has rarely been reported in PAH but is more frequently associated with PVOD / PCH and PAH associated with connective tissue diseases or congenital heart disease. Similarly, Raynaud's is usually associated with PAH associated with connective tissue disease but was also found to be a poor prognostic marker in "primary pulmonary hypertension" (202).

Clinical investigations required for the assessment of patients presenting with suspected idiopathic or heritable PAH, have three aims: 1) confirm a definitive diagnosis of PAH, 2) exclude other forms of pulmonary hypertension, especially associated PAH, 3) assess the severity of disease, which may aid risk stratification.

A definitive diagnosis of PAH requires right heart catheterisation to measure pulmonary artery pressures at rest and in a supine position. In addition, right heart catheterisation allows measurement of right atrial pressure (RAP), cardiac index (CI) and mixed venous oxygen saturations (S_vO_2) all of which are of prognostic value (1, 183, 202, 203). *BMPR2* variant carriers have been shown to have a younger age at disease onset and more severe pulmonary haemodynamic impairment (higher PVR and lower CI) compared to those with idiopathic PAH (30, 204).

A small subset of patients will have a positive response to an acute pulmonary vasodilator challenge. Identification of these patients is important as they will respond to long term calcium channel blockers (CCBs) and have a significantly better prognosis (205, 206). For vasoreactivity testing, patients inhale 40 parts per million (ppm) NO for 5 minutes prior to reassessment of their pulmonary haemodynamics (1, 207). An acute vasodilator response is defined as a drop in mPAP $\geq$ 10 mmHg to an absolute level < 40 mmHg and with no change in cardiac output (CO). *BMPR2* variant carriers are significantly less likely to have a vasoreactive response compared to those without a *BMPR2* variant (2.6 % vs. 16.2 %) (30, 208). Similarly, those in functional class 3 (4.7 %) and 4 (3.0 %) are less likely to respond to a vasodilator challenge compared to those in functional class 1 or 2 (9.8 %) (27).

Other investigations used to assess patients with pulmonary hypertension are summarised in Table 2.

Table 2. Clinical investigations used to assess patients with PAH

Investigation	**Variable / Feature**	**Utility**	**Prognostic in univariate tests**
Electrocardiogram (ECG)	Right axis deviation; right bundle branch block; right ventricular (RV) hypertrophy	Evidence of RV strain	
	Arrhythmias	Evidence of RV strain	(209)
Echocardiogram	Right atrial (RA) size; RV size, morphology and ejection fraction	Evidence of RV strain	
	Estimated systolic pulmonary artery pressure	Non-invasive assessment of pulmonary artery pressures	
	Pulmonary acceleration time		
	Tricuspid annular plane systolic excursion (TAPSE)	Evidence of RV strain	(210, 211)
	Left ventricular (LV) eccentricity index	Evidence of RV strain	(211)
	Pericardial effusion	Evidence of RV strain	(182, 211)
	Left atrial (LA) size; LV size, morphology and function; aortic and mitral valve morphology and function	Evidence of left heart disease contributing to pulmonary hypertension	
	Shunts	Exclude congenital heart disease	

Blood tests	N-terminal pro brain natriuretic peptide (NT-ProBNP)	RA stretch	(182, 203, 212)
	Autoimmune screen	Assess for features of connective tissue disease	PAH associated with connective tissue disease has a worse prognosis (182, 213)
	Full blood count	Polycythaemia may be associated with hypoxia or congenital heart disease	Red cell distribution width (214) Platelets (215)
	Renal function	Poor renal function may be associated with poor organ perfusion or use of diuretics	Creatinine (182, 216)
	C reactive protein	Marker of inflammation	(217)
	Iron indices	Low iron associated with PAH	Soluble transferrin receptor (218)
	Uric acid	Severity of disease	(219)
	Thyroid function	Associated with PAH	(129)
	Lipid profile	Assess cardiovascular risks	High density lipoprotein (220)

Lung function tests	Forced expiratory volume in 1 second / forced vital capacity (FEV_1 / FVC)	Evidence of parenchymal lung disease	FVC (221)
	K_{CO}	Low K_{CO} despite normal spirometry associated with PVOD / PCH	(182, 202)
Chest radiograph (CXR)	Enlarged hilar	In keeping with dilated pulmonary arteries	
	Globular heart	RV dilation ± pericardial effusion	
	Pleural effusion	More in keeping with PVOD / PCH	
Ventilation perfusion scintigraphy / single-photon emission computed tomography / lung perfusion magnetic resonance imaging (MRI)	Mismatched perfusion defects	Exclude CTEPH	
CT thorax	Pulmonary angiogram	Assess for presence of chronic thromboembolic material and arteriovenous malformations	

	Lung parenchyma	Centrilobular ground glass opacification, interlobular septal thickening and mediastinal lymphadenopathy associated with PVOD / PCH	Septal thickening and opacifications in Group 1 PAH (222, 223)
	RV size, pulmonary artery size	Assess severity of disease	
Cardiac MRI	RA size; RV size, function, ejection fraction and stroke volume	Assess severity of disease	(224, 225)
	Proximal pulmonary artery relative area change	Vascular stiffness	(225)
	Shunts	Exclude congenital heart disease	
	LV size, function and ejection fraction; valvular function	Exclude left heart disease or valvular heart disease	
Walk test (6-minute walk test or shuttle walk test)	Distance walked	Assess functional capacity	(182, 183, 203, 226-229)
Cardiopulmonary exercise test	Peak work; peak oxygen consumption (VO_2 max); anaerobic threshold; oxygen uptake efficiency slope	Assess functional capacity	

	Oxygen pulse; ventilatory equivalents; maximal heart rate; maximal blood pressure	Assess cardiac and ventilatory reserve	Maximal heart rate (227) VE/VCO_2 (227)
Right heart catheterisation	Systolic, diastolic and mean pulmonary artery pressure	Gold standard for confirming pulmonary hypertension	mPAP (202)
	PCWP	Surrogate for LA pressure – exclude significant left heart disease	
	RAP	Disease severity	(182, 183, 203, 221)
	CO and CI	Disease severity	(183, 202, 203)
	PVR	Disease severity	(182)
	S_vO_2	Disease severity	(203, 221)
	Vasoreactivity testing	Identify those who may benefit from CCBs	(221)
Left heart catheterisation	Left ventricular end-diastolic pressure (LVEDP)	Exclude significant left heart disease, when PCWP unreliable	
	Coronary angiogram	Exclude significant coronary artery disease	

Genetic counselling and clinical genetic testing

Current guidelines recommend that genetic testing be offered to patients with idiopathic, heritable and anorexigen associated PAH, as well as patients with PVOD / PCH (1). Genetic testing can be used to identify families at risk (screening of symptomatic index cases), identify at risk relatives (presymptomatic screening of relatives), aid family planning, determine pre-implantation genetic diagnosis and help risk stratify patients (230-232). It is thought that knowing your risk of carrying a disease associated variant and the risk to family members can be beneficial. Jones and Clayton showed that patients and relatives who undergo genetic testing reduced their distress levels more than those who declined further investigation (233). The most common reasons given for pursuing genetic testing was to determine the risk of passing a variant to children and the risk of carrying a variant for the individual (234, 235). Yet the uptake of genetic testing can be low. In the study by Jones et al. uptake was just 9 %. This may not be helped by the perception, amongst PAH healthcare professionals, that the utility of genetic testing and counselling is low (236).

Even prior to the recognition of the role of *BMPR2* in disease pathogenesis, microsatellite marker segregation was being used to identify at risk patients in families with heritable PAH. In the first such reported case, Morse and Barst identified an apparently unaffected sibling who carried a disease associated haplotype (segregating amongst affected family members) (237). Further assessment of the at-risk sibling uncovered that she did have heritable PAH and had been misdiagnosed with asthma.

Current PAH guidelines recommend that prior to clinical genetic testing patients and relatives receive genetic counselling to understand the nature of the tests and their implications (1). However, the implementation of genetic counselling and testing is usually in accordance with national guidelines (235). Even prior to genetic testing, genetic counsellors (and other trained healthcare professionals) can inform patients about the risks of carrying a disease-causing variant and the risk of developing disease. This is usually based on what is known about *BMPR2* variants, given that they account for the majority of heritable PAH cases. For example, an offspring of a parent who has a disease associated *BMPR2* variant has a 50 % chance of inheriting the variant. Yet the low penetrance of the variant means that a female child will only have a 21 % risk of developing PAH and a male child has just a 7 % risk. While the risk of

disease in a child, whose parent has idiopathic PAH and has not undergone genetic testing, is 4 % if the child is female and just 1 % if the child is male (assuming 17 % of those with idiopathic PAH have a variant in *BMPR2*). At risk individuals should be offered annual follow up with echocardiography to assess for the development of PAH (1).

Current practice for clinical genetic testing is first to screen for variants (single nucleotide variants (SNVs), indels and larger structural variants) in *BMPR2*; in those with idiopathic, heritable or anorexigen associated PAH, unless there are clear features of HHT or PVOD / PCH (1, 235). If no disease associated variants are identified then screening for variants in *ACVRL1*, *ENG* and *SMAD9* can be performed. If both these initial screens fail to identify a causal variant, then variants in the other genes associated with PAH can be assessed. Sanger sequencing and MLPA remain the gold standard for clinical genetic testing. Next generation sequencing (NGS) technologies allowing high throughput screening for variants in multiple genes may streamline this process (86, 238). NGS technologies are discussed further below.

Treatment

Treatment options for idiopathic and heritable pulmonary arterial hypertension remain limited. Current therapies licenced in the United Kingdom, Europe and the United States all act as pulmonary artery vasodilators. These licenced therapies do prolong survival, increasing exercise capacity and improve quality of life. However, there is currently no cure for the disease other than lung transplantation.

Prostaglandin I_2 (prostacyclin; normally synthesised in endothelial cells) and its analogues were first used to treat PAH (239, 240). They are agonists of the prostaglandin I_2 receptor (a G_s protein coupled receptor), and cause vasodilation. They can also activate other prostanoids receptors (241). Endothelial surface receptor binding leads to activation of adenylyl cyclase, increasing cyclic adenosine monophosphate (cAMP), which in turn leads to the activation of protein kinase A (PKA) and ultimately phosphorylation of the myosin light chain kinase. Activation of other prostanoids receptors on pulmonary artery smooth muscle cells may cause anti-proliferative effects. Prostacyclin analogues currently used include, epoprostenol (intravenous), iloprost (intravenous and inhaled), treprostinil (intravenous, subcutaneous, and inhaled), beraprost (oral) and the prostaglandin I_2 receptor agonist, selexipag (oral).

Randomised control trials have shown a prognostic benefit with intravenous epoprostenol therapy (226).

Endothelial derived NO is another potent pulmonary artery vasodilator and acts by activating soluble guanylate cyclase, thereby increasing cytosolic cyclic guanosine monophosphate (cGMP) in pulmonary artery smooth muscle cells (242). cGMP causes a reduction in cellular calcium and hyperpolarisation of the cell membrane by activating potassium channels. Phosphodiesterase 5 inhibitors (PDE5i) such as sildenafil and tadalafil prevent the degradation of cGMP enhancing smooth muscle relaxation. Sildenafil monotherapy has been shown in randomised control trials to improve functional capacity and quality of life (243, 244).

Endothelin 1, produced by endothelial cells, is a potent vasoconstrictor which exerts its effects through endothelin receptors A and B (ET-A and ET-B). ET-A and ET-B are Gq coupled receptors. Activation of the receptors leads to an increase in inositol triphosphate (IP_3) and diacylgycerol (DAG) by phospholipase C. IP_3 binds to ligand gated calcium channels to release calcium into the cytoplasm and trigger smooth muscle contraction. Whereas DAG activates protein kinase C ultimately leading to increased phosphorylation of myosin light chain by inhibition of myosin light chain phosphatase. Endothelin receptor antagonists (ERAs), such as bosentan, ambrisentan and macitentan, block this signalling pathway leading to vasodilation. Bosentan and ambrisentan monotherapy has been shown to improve exercise capacity and symptoms, while also improving time to clinical worsening in extended studies (245-248).

In more recent clinical trials, owing to the fact it would be unethical to run placebo-controlled trials on treatment naïve patients, randomised clinical trials have assessed the impact of introducing an additional drug class on top of existing therapies (249, 250). These studies showed that the addition of a new class of drug improves composite end-points, such as time to clinical worsening (that includes death, transplantation, initiation of intravenous therapy or worsening of pulmonary hypertension). A goal directed approach to the treatment of pulmonary hypertension has supported sequential combination therapy or increasing the dose of existing therapies (1, 251, 252). More recently, studies have shown that upfront combination therapy with two classes of drug (PDE5i and ERA) results in improved time to

clinical worsening (253, 254). A single French trial showed a benefit from upfront combination of all 3 classes of currently licenced pulmonary artery vasodilator therapies in those with severe disease (255).

CCBs, such as diltiazem and nifedipine, are indicated only in those who show a vasoreactive response to an acute pulmonary vasodilator challenge. The dose of CCBs should be titrated up until the vasoreactivity is lost on follow up testing or patients do not tolerate further dose escalation. Failure to normalise pulmonary artery pressures and improve functional class should lead to commencement of pulmonary artery vasodilator therapies (1).

Atrial septostomy creates a right atrial to left atrial shunt to improve pulmonary haemodynamics and improve left ventricular filling pressures (256, 257). It is now rarely performed and current guidelines limits its role to a palliative therapy or as a bridge to lung transplantation (1, 258).

Lung transplantation remains the only potential cure for PAH and is considered in those with an inadequate haemodynamic and/or functional improvement following medical therapy (1). The mortality amongst patients with idiopathic PAH on lung transplantation waiting lists is greater than those with other chronic lung diseases. Therefore, prompt and appropriate referral for lung transplantation assessment is required (1, 259-261). On the other hand, with improvements in medical therapies, the rates of lung transplantation for PAH have fallen (262). Double lung transplants are the most common form of transplant offered to idiopathic PAH patients as the RV has been shown to reverse remodel and improve function following the reduction in RV afterload accompanying lung transplantation (263-265).

Prognosis

The median survival of patients with idiopathic and heritable PAH prior to the advent of pulmonary artery vasodilator therapies from the US NIH Study was just 2.8 years; while 1, 3 and 5-year survival was 68 %, 48 %, and 34 % respectively (202). The changing demographics of the disease and differences in cohort composition can confound comparison to this early study, but in general survival in comparable and more recent cohorts have improved (Table 3). In the UK National Registry 1, 3, and 5-year survival are reported as 93 %, 73 % and 61 %

respectively (25). Following lung transplantation the median survival in patients with idiopathic PAH is 10 years, although there is a significant early mortality risk for these patients (265).

An important observation in many of these studies is the worse prognosis of incident patients when compared to patients with prevalent disease (202, 266, 267). Therefore, survival analyses from registry data that include both prevalent and incident cases should account for survivor bias. This bias arises from the fact that prevalent patients have already survived a certain period of time prior to enrolment in the study. Therefore, they can be considered to have either less severe disease or a better response to treatments compared to patients diagnosed at a similar time point but who die prior to enrolment in the study. In contrast, incident patients consist of both patients with a poor prognosis who die early and patients with a good prognosis. Consequently, incident patients will have a worse overall prognosis compared to prevalent patients. This may confound the results of survival analyses trying to identify other prognostic factors.

A family history of PAH was an independent marker of poor prognosis in a cohort of patients with PAH in comparison to patients with idiopathic PAH (182, 268). In keeping with this finding Evans et al. report that patients with *BMPR2* variants have a significantly worse prognosis than patients with idiopathic PAH (30). This was a consequence of poor RV function in the former group. Other variables consistently reported as independent prognostic indicators in multivariate models include age, gender, functional class, 6mwt distance, NT-ProBNP, pericardial effusion, D_{LCO}, RAP, CO, CI and PVR (182, 183, 269-272). Although older patients have less severe pulmonary haemodynamics they have a worse all-cause mortality (273). Although, coronary artery disease is not an independent predictor of survival (274). With regard to gender, male patients aged over 60 years at diagnosis have the worst prognosis, whereas no difference in prognosis is seen between male and female patients under the age of 60 years (275).

Table 3. Summary of studies reporting survival in patients with idiopathic and heritable PAH

Study	**Cohort**	**1 / 2 / 3 / 5-year survival (%)**	**Median survival (years)**	**Significant prognostic variables on multivariate analysis**
National Institute of Health (US) (202)	1981 – 1985 Incident and prevalent "Primary pulmonary hypertension" PCWP < 12 mmHg	68 / *NA* / 48 / 34	Overall 2.8 Incident 2.6 Prevalent 3.2	Functional class 3 or 4, mPAP, RAP, CI, diffusion coefficient for carbon monoxide (D_LCO),
Mexican cohort (221)	1977 – 1991 "Primary pulmonary hypertension" mPAP ≤ 22 mmHg		Combined 4.0 No vasodilator therapy 2.1 Vasodilator therapy 5.0	FVC, CI, RAP
Pulmonary Hypertension Connection Registry (US) (269)	1991 – 2007 Incident Idiopathic, heritable, anorexigen associated PAH	91 / *NA* / 75 /65		Age, functional class, RAP, CI
Kawut et al. (US single centre) (195)	1994 – 2002 Incident Idiopathic, heritable or anorexigen associated PAH	87 / *NA* / 75 / 61		Ethnicity, albumin, CI, vasoreactivity, warfarin use

Hannover Cohort (Germany) (271)	1995 – 2005 Incident and prevalent Idiopathic PAH	93 / 77 / 67 / 60		6mwt distance, RAP, P_aCO_2
Spanish Registry of Pulmonary Arterial Hypertension (276)	1998 – 2008 Incident and prevalent Idiopathic PAH	89 / *NA* / 77 / 68		
Ogawa et al. (Japan) (277)	1998 – 2012 Incident and prevalent Idiopathic and heritable PAH	98 / 96 / 96 /96		
Jing et al. (China) (192)	1999 – 2004 Idiopathic and heritable PAH	68 / 57 / 39 / 21		
Jansa et al. (Czech Republic) (278)	2000 – 2007 Incident and prevalent Idiopathic and heritable PAH	85 / 70 / 62 / *NA*		Incident group only: Creatinine clearance, 6mwt distance
REVEAL (US) (279)	2000 – 2009 Incident and prevalent Idiopathic and heritable PAH	91 / *NA* / 74 / 65		
Pulmonary Hypertension Registry of the	2001 – 2009 Incident	93 / 84 / 73 / 61		

United Kingdom and Ireland (25)	Idiopathic, heritable and anorexigen associated PAH			
French Pulmonary Hypertension Network (183, 266)	2002 – 2003 Incident and prevalent Idiopathic, heritable and anorexigen associated PAH	Incident: 89 / 68 / 55 / *NA* Prevalent: 89 / 77 / 69 / *NA*		Gender, 6mwt distance, CO
Zeng et al. (China) (272)	2006 – 2009 Incident Idiopathic PAH	84 / 74 / 71 / *NA*		Age, BMI, pericardial effusion, absence of pulmonary artery vasodilator therapy
Zhang et al. (China) (280)	2007 – 2009 Incident Idiopathic PAH mPAP > 30mmHg during exercise included	92 / 80 / 75 / *NA*		Including PAH associated with connective tissue disease: D_LCO, functional class

Pulmonary veno-occlusive disease and pulmonary capillary haemangiomatosis

Epidemiology

PVOD / PCH are extremely rare forms of PAH. The French registry data contains arguably the most thorough assessment of PVOD / PCH in any national registry. However, their population includes at risk groups, such as North African families with high levels of consanguinity that predispose them to autosomal recessive genetic conditions such as PVOD / PCH. Therefore, the French registry data may not be generalisable to other populations. It is estimated that the prevalence of PVOD / PCH is 1 to 2 per million of the population (281). Whilst, the UK 2015 National Pulmonary Hypertension Audit states that 1 % of all referrals in 2013 to 2014 (approximately 20 patients) resulted in a diagnosis of PVOD / PCH (28).

Pathophysiology of PVOD / PCH

Genetic basis

The first confirmed familial cases of PVOD / PCH were reported by Voordes et al. in 1978 (282). In this family two brothers from non-consanguineous parents died at 8 weeks and 3 months from PVOD. The authors concluded that the rarity of familial PVOD may suggest that environmental rather than genetic factors were the cause of the disease. A few years later a recessive pattern of inheritance was postulated for familial PCH in a report describing three siblings with PCH (283).

In their seminal paper, Eyries et al. studied French families with PVOD (20). They noted that the disease was seen most commonly in children from consanguineous parents, suggesting an autosomal recessive pattern of inheritance. Through WES and Sanger sequencing they were able to identify biallelic *EIF2AK4* variants as the cause of familial PVOD in all 13 families studied. They went on to show that approximately 25 % of patients with apparently sporadic PVOD also carried biallelic *EIF2AK4* variants.

Soon afterwards, Best et al. reported biallelic *EIF2AK4* variants in patients with PCH, confirming that PVOD and PCH are part of the same disease spectrum (21). In this study one family with PCH was initially screened with WES to identify the causal variant. Subsequent assessment of patients with sporadic PCH revealed 20 % also carried biallelic *EIF2AK4*

variants. Interestingly, they also report a second family with an apparent autosomal dominant pattern of inheritance that did not carry *EIF2AK4* variants.

As discussed earlier, two recent reports have described a single family and a large kindred of Iberian gypsies with apparent heritable PAH, all of whom also carried biallelic *EIF2AK4* variants (22, 93). However, the phenotypic characterisation of patients in both studies was insufficient.

Variants in *BMPR2* have also been reported in 6 patients with apparent PVOD (41, 281, 284, 285). However, some authors have argued that this is likely to be due to misclassification of these patients (281).

GCN2 is part of the integrated stress response that stops global protein translation in response to environmental stressors such as amino acid starvation (286). Amino acid starvation results in an increase in uncharged transfer RNAs (tRNAs). Uncharged tRNAs, that would normally attach to essential amino acids that cannot be synthesised, bind and activate GCN2 (287). Therefore, GCN2 can act as a sensor for amino acid paucity and starts a series of reactions preventing the transfer of methionyl-initiator tRNA to the 40s ribosomal complex that is required to initiate translation.

Activated GCN2 phosphorylates serine 51 of the eukaryotic translation initiation factor 2 alpha subunit (eIF2α) (288, 289). eIF2α binds guanosine-5'-triphosphate (GTP) / guanosine-5'-diphosphate (GDP) and is part of the heterotrimeric complex eIF2. Prior to protein translation, EIF2 binds with the guanine nucleotide exchange factor (eIF2B) to convert GDP to GTP. This conversion of GDP to GTP is inhibited if the serine 51 residue of eIF2α is phosphorylated (290). eIF2-GDP is unable to bind methionyl-initiator tRNA preventing the initiation of translation.

Conversely, the sparsity of primed 40s ribosomal complexes facilitates the translation of activating transcription factor 4 (ATF4), which switches on pathways related to the integrated stress response. The translation of ATF4 is usually reduced in amino acid replete conditions by the presence of inhibitory upstream open reading frames (291). When there is a scarcity

of amino acids and few primed 40s ribosomal complexes, translation initiation is improved at the ATF4 start codon.

The mechanism by which disease associated variants in *EIF2AK4* (that are unable to phosphorylate eIF2α in response to cellular stresses) leads to PVOD / PCH is unknown. However, the importance of its role is suggested by the fact that environmental exposures associated with the development of PVOD / PCH also cause a reduction in GCN2 protein levels (292).

Environmental exposures and toxins

Alkylating agents

Registry data from France has linked exposure to alkylating chemotherapy drugs such as cyclophosphamide and mitomycin-C, to the development of PVOD / PCH, giving weight to previous case reports suggesting an association (293). Furthermore, animals exposed to these agents developed pulmonary hypertension and demonstrated pulmonary venous remodelling (292, 293). Perros et al. linked exposure of mitomycin-C to GCN2 by demonstrating that administration to rats resulted in a reduction of GCN2 in lung tissue.

Organic solvents

A case-control study of consecutive patients presenting to the French National PH Centre showed an increased risk of exposure to organic solvents, particularly trichloroethylene, amongst patients with PVOD / PCH compared to PAH (294). In the study Montani et al. reported that PVOD patients with biallelic *EIF2AK4* variants were younger at diagnosis compared to patients exposed to organic solvents.

Histological descriptions

In their 1970 paper of 156 patients with a clinical diagnosis of “primary pulmonary hypertension” Wagenvoort and Wagenvoort were able to identify 5 patients (3 %) who they classified histologically as PVOD (105). These patients had early onset disease (average age just 16 years) and were characterised by severe obstruction of the small pulmonary veins and

venules. Although pulmonary arterial changes were also observed this was not as severe compared to the “vasoconstrictive primary pulmonary hypertension” group.

Pulmonary venous remodelling is the hallmark of PVOD / PCH (281, 295). Assessment of the lung parenchyma, often post-mortem or following lung transplantation, remains the gold standard for the diagnosis of PVOD / PCH. However, the histological features of PVOD / PCH can be heterogeneous in distribution, requiring careful assessment of all lobes of the explanted lungs for an accurate histological diagnosis (109).

The main histological features of PVOD / PCH are intimal fibrosis of the paraseptal veins and capillary haemangiomatosis. Intimal fibrosis and thrombosis of the pulmonary venules and paraseptal veins can lead to occlusion of these vessels. The media of the paraseptal veins also show signs of “arterialisation” with medial hypertrophy and increased deposition of elastic fibres. In PVOD / PCH, unlike other forms of PAH (such as connective tissue disease associated PAH and PH associated with mitral stenosis), these pulmonary venous changes are usually florid and the predominant histological feature (296, 297).

Haemangiomatosis of the pulmonary capillaries, whereby there is invasion of the pulmonary capillaries into surrounding tissue, is another hallmark of the disease. It is found in both PVOD and PCH, suggesting the two are part of the same disease spectrum (295). Pulmonary arteriopathy may also be present but is not the predominant feature. Other histological features associated with the disease include dilation of pulmonary lymphatics, presumably due to increased interstitial oedema, and enlarged mediastinal lymph nodes. Hemosiderin laden alveolar macrophages are also characteristic of PVOD / PCH and can be detected in bronchoalvelolar lavage fluid (298). Yet, these histological features do not always correlate with the clinical diagnosis. Stacher et al. reported two patients with a clinical diagnosis of PVOD who had histological features more in keeping with those with idiopathic PAH (106). Autopsy series have also confirmed that approximately 10 % of patients with a clinical diagnosis of PAH may have a histological diagnosis of PVOD (299).

Demographics

The age of disease onset for PVOD / PCH is extremely wide from infancy to old age (20, 282, 300). Interestingly, a bimodal distribution for age of diagnosis was suggested by Montani et al. with patients with biallelic *EIF2AK4* variants presenting at an earlier age compared to those with exposure to organic solvents (294).

The gender of patients also appears to vary depending on disease aetiology. Those with biallelic *EIF2AK4* variants have an equal gender split, whereas there is a male predominance amongst those with exposure to organic solvents, likely reflecting gender differences in occupation (294).

Clinical features and assessments

The clinical features and results of clinical assessments for PVOD / PCH are not specific and similar to PAH. Autopsy series have suggested that up to 10 % of patients with histological evidence of PVOD / PCH may be misdiagnosed as PAH based on clinical assessments. Therefore, histological confirmation remains the gold standard for diagnosis. However, lung biopsy in patients with all forms of PH is associated with morbidity due to bleeding (1). Consequently, it is not routinely performed.

Although not specific for PVOD / PCH, as it is also reported in PAH, digital clubbing has been described as a feature of the disease (301, 302). The most striking clinical characteristics of PVOD / PCH is a lower arterial oxygen saturation and more marked desaturation on exertion compared to patients with PAH (300). This is associated with a reduced K_{CO} in patients with PVOD / PCH (303). Godinas et al. assessed the diffusion capacity of the lung for both carbon monoxide (D_{LCO}) and nitric oxide (D_{LNO}) to determine whether this reduction in gas exchange was due to reduced blood flow or impaired diffusion across the alveolar membrane. They report that patients with PVOD had a higher D_{LNO} / D_{LCO} ratio compared to patients with PAH, suggesting that the reduction in gas exchange was due to reduce blood flow through the capillary bed (304). Haemoptysis, which is also associated with *BMPR2* variants, is not more common amongst patients with PVOD / PCH despite the presence of hemosiderin laden macrophages in bronchoalvelolar fluid (112, 298, 300).

Apart from confirming the presence of pulmonary hypertension, right heart catheterisation does not provide any additional information to suggest a diagnosis of PVOD / PCH. It is generally accepted that most patients with PVOD / PCH do not have an acute vasoreactive response, but some patients (even with biallelic *EIF2AK4* variants) do show a positive response (305, 306). Some have argued that a response to inhaled nitric oxide can be used as a marker for response to pulmonary artery vasodilator therapies, although this has not been systematically investigated (307). The PCWP for patients with PVOD / PCH is usually less than 15 mmHg unlike patients with PH associated with left heart disease.

High resolution CT of the lung parenchyma is a useful tool for assessing for PVOD / PCH. Radiological features such as mediastinal lymphadenopathy, interlobular septal thickening, and centrilobular groundglass opacification may be suggestive of PVOD / PCH but all these features can also be found in other forms of PAH (222, 308). The presence of 2 of these radiographic features (along with a K_{CO} < 60 % predicted, mPAP $\geq$ 25 mmHg and PCWP $\leq$ 15 mmHg) have been used in studies to identify patients highly likely to have PVOD / PCH (306).

Treatment and prognosis

No large randomised control trials have been performed in patients with PVOD / PCH. Case reports and series have suggested sildenafil and intravenous prostanoids therapies are of benefit in some patients with PVOD / PCH (309, 310). However, these pulmonary artery vasodilator therapies can result in fatal pulmonary oedema and therefore should be used with caution (300, 310-313). No trials have assessed which patients may be at highest risk of developing pulmonary oedema. Montani et al. report that pulmonary oedema can occur in both *EIF2AK4* variant carriers (21 %) and those with idiopathic PVOD / PCH (23 %) (306). Consequently, lung transplantation remains the only form of treatment for some patients. Early referral for lung transplantation assessment is recommended as it is reported patients with PVOD / PCH have a higher risk of death while on lung transplantation waiting lists compared to patients with PAH (314).

The prognosis for patients with PVOD / PCH is poor in comparison to PAH (276). Amongst patients with biallelic *EIF2AK4* variants the 1, 2 and 3-year event free survival rates were 63

%, 52 % and 32 %. The rates for those without *EIF2AK4* variants were similar: 75 %, 44 % and 34 % (306).

UK Pulmonary Hypertension Centres

The clinical care of patients in the UK with PAH is centralised in 8 National Pulmonary Hypertension Centres and their satellite units (Appendix 1). The centralised system of care has resulted in specialist centres with a wealth of clinical experience and the infrastructure to conduct research into the disease. It facilitates the recruitment of patients in large numbers and standardises the work up and management of patients through the national pulmonary hypertension audits (28).

National Institute of Health Research BioResource – Rare Diseases Study

The National Institute of Health Research (NIHR) BioResource – Rare Diseases (BRIDGE) Study (https://bioresource.nihr.ac.uk/rare-diseases/rare-diseases/) was set up to investigate the genetic basis of rare inherited diseases whose pathogenesis remained uncertain. This was to be achieved through next generation whole genome sequencing (WGS), conducted by Illumina Inc (USA), and deep phenotyping. Idiopathic PAH was chosen as one of the rare diseases suitable for further evaluation. The other rare disease cohorts assessed as part of the study are listed in Appendix 2.

Next generation sequencing

Benefits and drawbacks of next generation sequencing

NGS technologies differ from conventional Sanger sequencing and capillary electrophoresis by providing high-throughput, low cost, massively parallel (adapters used to create DNA libraries can be coded to identify different samples) and unbiased sequencing capabilities. NGS technologies can sequence the entire genome (WGS), protein coding sequences (WES) or just a panel of pre-defined genes (315). However, NGS technologies also have disadvantages compared to traditional Sanger sequencing. For example, the read lengths generated by NGS are comparatively short compared to that generated by capillary electrophoresis, this can make sequence alignment to a reference genome more difficult. There is also a higher error rate in base calling, which may lead to false positive variant calls.

In an assessment of a particular type of second generation sequencing technology called sequencing by synthesis (SBS; Illumina, USA), Bentley et al. compared SNVs called by SBS to genotyped calls and report > 99.5 % concordance between the two. Although, the false positive rate with SBS was 2.5 % (316). Sanger sequencing is recommended to confirm a NGS called variant (317). Furthermore, in their study of patients with heritable PVOD / PCH Eyries et al. demonstrated variants in *EIF2AK4* that were initially not called by WES but were subsequently identified by Sanger sequencing (20). This lack of sensitivity is more significant in WES compared to WGS due to the biases introduced by the need for exon specific primers with WES.

Another disadvantage of NGS, and in particular WGS, is the sheer volume of data generated. Consequently, false positive results and incidental findings need to be assessed and acted on. NGS based gene panels overcome this issue as only a small, targeted number of genes are assessed. This can be a useful first clinical screen for diseases where the genetic basis is well understood. Nonetheless, it is limited and biased as a research tool for identifying variants in novel disease associated genes or genetic modifiers. Even in complex polygenic diseases rare variation with relatively large affect sizes may explain the missing heritability seen in large studies assessing the role of common variation in disease (318). Therefore, WES and WGS are well suited to further investigate both rare and common genetic disorders (319). WES sequencing technology has already been used in families with heritable PAH and PVOD / PCH to identify variants in *EIF2AK4*, *CAV1* and *KCNK3* (20, 21, 75, 80). Importantly, all these studies confirmed identified variants through Sanger sequencing.

While it is estimated 85 % of disease causing variants are in the protein coding sequence, WES has been shown to be less accurate and more biased compared to WGS (320). WES requires exon specific primers and polymerase chain reaction (PCR) amplification of these protein coding sequences. Whereas with the latest library preparation kits no PCR amplification is required for WGS, reducing the biases associated with primer selection. The use of primers can also result in a reduction in read depth (the number of sequencing strands covering a single position in the genome) in areas of the genome where there is an abundance of repetitive guanine and cytosine residues (GC rich areas) (321). WGS is also a better tool to assess structural variation across the entire genome by analysing drops in read depth and

differences in the distance between pair-end reads. Even in the protein coding sequence, WGS is more sensitive at identifying variants compared to WES (322). On the other hand, WES generally provides higher read depths, which may make it more suitable for identifying somatic variants and mosaicism compared to WGS.

Sequencing by synthesis

Illumina's SBS technology is currently the most commonly used second generation sequencing technology (316, 317, 323). The first step in SBS is the creation of a library of DNA fragments. Initially, genomic DNA or complementary DNA (cDNA) molecules are randomly fragmented and ligated at both the 3' and 5' ends with adapter sequences. Fragments of a specific size are taken forward. These fragments are amplified by PCR to create a DNA library. Although the latest WGS technologies do not require PCR amplification.

In preparation for sequencing, the library is exposed to a flow cell surface lined with complementary oligonucleotides to which the adapter sequences bind. The flow cell surface is where the sequencing reactions occur. However, the fluorescence signal generated by a single molecule is too weak to be picked up. Therefore, the bound DNA fragments undergo a process of bridge amplification, through solid state PCR, to create a cluster of approximately a million identical fragments in close proximity with each other. It is the fluorescent signals generated by these clusters that are detected by the sequencer.

During each sequencing cycle, all four dideoxynucleotides (ddNTPs; adenine, guanine, cytosine and thymine) are flowed across the flow cell simultaneously. The ddNTPs competitively bind complementary nucleotides in the DNA fragments attached to the flow cell surface. Each added ddNTP is modified on the 3' end to carry a specific fluorophore and blocking group. This allows identification of the ddNTP once it has been incorporated and prevents further binding to the 3' end, ensuring only one ddNTP binds each DNA fragment during each cycle. An advantage of SBS over other NGS technologies is the presence of all four ddNTPs during each sequencing cycle (316). The competition for binding between the ddNTPs enhances the accuracy with which the ddNTPs bind the appropriate complementary nucleotides on the attached DNA fragments.

At the end of each reaction the fluorophores emit a flash of light after being excited by a laser. The sequencer determines the nucleotide incorporated within each cluster by identifying the wavelength of the light emission. This entire process is repeated after the fluorophores and blocking groups are removed, restoring the 3' ends of the incorporated ddNTPs to allow further extension. In this manner, the sequence of the DNA fragments in each clonal cluster is determined one nucleotide at a time. The number of times the process is repeated determines the length of the sequence (the read length).

For paired-end sequencing, double stranded DNA is created from the DNA fragments and the original strand removed from the flow cell before the process is repeated for the new complimentary strand. As the DNA fragments are of a known length and the read length is also known, the distance between the two paired-end sequences can be determined. This enhances alignment of reads to highly repetitive regions of the reference genome (mitigating one of the drawbacks of NGS technology) and facilitates the identification of structural variants (both small indels and larger deletions and insertions) (324). Paired-end sequencing also aids the identification and removal of PCR duplicates and has been shown to identify more SNVs compared to single-read sequencing.

Bioinformatics pipeline

Following sequencing several quality control (QC) and processing steps are required prior to the identification of both SNVs and structural variants. There is no gold standard but four steps are required: 1) QC of the reads, 2) alignment of reads against a reference genome, 3) variant and genotype calling and 4) variant annotation (325). Multiple open source and proprietary software options exist for each of these steps.

As well as generating the reads, the sequencer can also estimate the quality of base calls at each position of a read (Phred quality score). This is based on numerous factors including signal intensity, signals from surrounding clusters and the location in a read (less accuracy at longer read lengths and at the very beginning of a read). This information along with the base calls, information regarding the sequencer and flow cell are contained in FASTQ text files generated by the sequencer.

Further QC checks are performed on these FASTQ files. They can be analysed to assess Phred quality scores relative to read position, GC content, and the percentage of each nucleotide relative to the position in a read. These provide some estimates of the quality of the reads. After removal of the adapter sequences from the reads, the reads are further trimmed to remove inaccuracies at the beginning and end of each read. Furthermore, reads with extreme GC content can be removed to increase the overall quality of the remaining reads.

The next step is to align and sort the reads against a reference genome. The Genome Reference Consortium (GRC) curates reference sequences for humans and other organisms. The current iteration for humans (GRCh38) was released in December 2013 and consists of sequencing data from many individuals. This superseded GRCh37, which was released in 2009. GRCh38 increases the number protein coding sequences and improves sequence representation so that fewer reads from NGS remain unaligned (326). A particular issue with NGS generated reads is their comparatively short length, which means that it is harder to align to a reference genome, especially in areas of high variability (e.g. the major histocompatibility complex) or areas with many repetitive sequences areas (e.g. areas of high GC content) (327). Alignment software uses algorithms to maximise the consensus between a read and a section of the reference genome (328). However, variation in the reads, which can either be real SNVs or sequencing errors, complicate this process. Alignment software can take into account the quality scores of bases to identify potential errors in the reads. Ultimately unaligned reads or reads that align with more than one part of the genome are removed. The output of alignment software is usually in the Sequence Alignment/Map (SAM) format or its binary equivalent, BAM.

Variant and genotype calling software takes aligned and sorted sequence information and determines differences compared to the reference genome. PCR duplicates must be removed at this stage to avoid potential variation in the duplicates biasing variant calling. The software uses Bayesian statistical models to determine if there is difference at a particular position compared to the reference genome (329). Several factors are taken into consideration depending on the software, such as the quality of the base call, the position within a read, the number of reads that call a variant, and variation around the position of interest. While databases of SNPs such as dbSNP can be used to determine prior probabilities for a variant at

a particular position. SNVs that fail checks such as deviation from the Hardy-Weinberg equilibrium, have insufficient coverage or poor-quality base calls are excluded to reduce the rate of false positive variants called. A variant call format (VCF) file contains information about variation in a genome and is the standard output of variant and genotype calling software. VCF files typically also contains information about where and what the variant is, a quality score for the variants and a summary of filters that a variant has passed. Additional information such as allele frequencies, deleteriousness scores, and associated phenotypes can be added.

Once a variant has been identified further information is required to determine its significance and this is referred to as variant annotation. Variant annotation provides information of the consequence of a variant on transcription and translation (e.g. stop gained or missense variant), the frequency of the variant in previously studied populations, assessment of its conservation between different species and *in-silico* predictions of its deleteriousness. Databases, such as the Exome Aggregation Consortium (ExAC) database and the UK10K database, can be used to assess the allele frequency of variants in specific populations (330, 331). There are several different bioinformatic tools that predict the deleteriousness of a variant. Their methods vary but are generally based on *in-silico* predictions of biochemical affects, frequency in the general population and conservation between species. Examples include the combined annotation dependent depletion score (CADD score), PolyPhen-2 prediction and SIFT prediction (332-334).

Medical Research Council / British Heart Foundation National Cohort Study of Idiopathic and Heritable Pulmonary Arterial Hypertension

The Medical Research Council (MRC) / British Heart Foundation (BHF) National Cohort Study of Idiopathic and Heritable Pulmonary Arterial Hypertension (Cohort Study) was closely associated with the NIHR BRIDGE PAH Study. The aims of the study were to longitudinally follow up patients with PAH and PVOD / PCH in the UK, biobank samples from patients, recruit unaffected family members for longitudinal follow up and conduct epidemiological research.

Hypothesis

My hypothesis was that idiopathic and heritable PAH are heterogeneous disorders, with diverse aetiologies, clinical features and outcomes, and that WGS and careful phenotyping of these patients could help identify clinically important subgroups.

Aims

My aims were to:

1) To phenotype a contemporary cohort of patients with idiopathic and heritable PAH, as well as PVOD / PCH, from diagnostic data collected through the NIHR BRIDGE PAH Study.
2) To identify prognostically important characteristics in patients with idiopathic PAH and use this to identify clinically important subgroups.
3) To assess the phenotypes of patients with variants in the known disease-associated genes, in particular those with *BMPR2* variants.
4) To assess the phenotypic consequences of biallelic *EIF2AK4* variants.

Methods

My Thesis is the result of two highly collaborative studies (the NIHR BioResource – Rare Diseases Study [NIHR BRIDGE Study] and the MRC / BHF National Cohort Study of Idiopathic and Heritable Pulmonary Arterial Hypertension [Cohort Study]). My direct contributions to the Thesis are highlighted in bold in the Methods Section below.

NIHR BioResource – Rare Diseases Study

The NIHR BRIDGE Study, was set up to investigate the genetic basis of rare heritable diseases whose pathogenesis has not been fully elucidated. The study is comprised of several disease cohorts of which PAH was one (Appendix 2). The study involved WGS, performed by Illumina Inc (USA), and deep phenotyping of the disease cohorts. Patients recruited to other disease cohorts acted as controls for the PAH sub-study. Recruitment to the NIHR BRIDGE PAH Study started in January 2013 and was closed in May 2017 after recruiting the allocated number of patients (1250).

MRC / BHF National Cohort Study of Idiopathic and Heritable Pulmonary Arterial Hypertension

The Cohort Study is a longitudinal study recruiting UK patients with PAH and their relatives. Patients are followed up at approximately 6 monthly intervals as part of their routine clinical care. Clinical information from these visits are prospectively captured for the study. In addition, blood and urine samples from these patients are biobanked for future studies. Patients have the choice to take part in several sub-studies including an epidemiology study and a MRI study. Recruitment to the Cohort Study commenced in April 2013 and the study is still open to recruitment.

My PhD started after the NIHR BRIDGE Study and the Cohort Study commenced patient recruitment. **I was involved with later amendments to the studies and was responsible for the overhaul of the phenotype data captured for the studies as specified below.**

Ethical approval and consent

Ethical approval for the NIHR BRIDGE Study was organised by the NIHR BRIDGE study team for all disease cohorts. All patients recruited to the NIHR BRIDGE Study provided written informed consent by signing either NIHR BioResource – Rare Diseases consent forms (Research Ethics Committee ID: 13/EE/0325) or local tissue bank consent forms that specified the use of biobanked blood samples for genetic sequencing and capture of clinical data (including paper records, electronic records and imaging studies). Prospectively recruited patients from the UK signed NIHR BioResource – Rare Diseases consent forms. Deceased UK patients recruited through local tissue banks and non-UK patients had previously signed local tissue bank consent forms.

Patients from the UK were recruited to the Cohort Study for longitudinal follow up. They provided written informed consent by signing MRC / BHF National Cohort Study of Idiopathic and Heritable Pulmonary Arterial Hypertension consent forms (13/EE/0203). Ethical approval for the Cohort Study was organised by the PAH coordinating team *[Carmen Treacy, Katherine Yates, Jennifer Martin]* and the principle investigator, Prof. Nicholas Morrell. **I helped draft study amendments, including specifying the recruitment of patients with PVOD / PCH for both studies and control subjects for the Cohort Study.**

Explanted lung tissue from a single patient with PAH was obtained through the Papworth Hospital Research Tissue Bank (08/H0304/56). Patients recruited to the tissue bank provided written informed consent for use of explanted organs and tissue for research, including genetic studies.

Patient recruitment

For the NIHR BRIDGE Study, patients were recruited from all the UK pulmonary hypertension specialist centres as well as specialist centres in Europe (Appendix 1). Research nurses and coordinators at each centre were responsible for patient recruitment, initial blood sample processing and data entry.

Patients were recruited to the study from specialist pulmonary hypertension centres if they had a diagnosis of idiopathic, heritable or drug / toxin associated PAH based on the latest clinical guidelines (1). Patients with PVOD / PCH were also recruited. The diagnosis was based on right heart catheterisation demonstrating a mPAP $\geq$ 25 mmHg and PCWP $\leq$ 15 mmHg measured at rest and in a supine position. For a diagnosis of idiopathic, heritable or drug / toxin associated PAH, other forms of PH needed to be excluded. For a diagnosis of PVOD / PCH a clinical suspicion of the disease was required based on computed tomographic imaging of the lung parenchyma in the absence of a histological diagnosis.

Throughout this Thesis, patients are classified as idiopathic PAH or familial PAH based on the absence or presence of a family history of the disease. The term heritable PAH does not distinguish between PAH patients with variants in disease associated genes (with or without a family history of the disease), and patients with a family history of PAH with no identifiable variants in disease associated genes. In this Thesis the term heritable PAH is used intentionally to include those with an identifiable disease associated variant and / or a family history of the disease. It is also used when referring to previous publications and guidelines.

Inclusion criteria

1) Incident and prevalent patients with a diagnosis of:
    a) Idiopathic PAH
    b) Heritable PAH
    c) Drug and toxin associated PAH
    d) PVOD
    e) PCH
2) Patients able to provide written informed consent

Exclusion criteria

1) Patients diagnosed with other forms of pulmonary hypertension
2) Patients unable to provide written informed consent

Phenotype data

OpenClinica

Phenotype data was captured using an open-source secure online tool called OpenClinica© (OpenClinica LLC, USA). Electronic case report forms (eCRFs) were adapted to capture demographic information and clinical assessments from both the time of diagnosis (Cohort and NIHR BRIDGE Studies) and from follow up visits (Cohort Study). The eCRFs were created on Microsoft Excel (Microsoft, USA) templates, which were uploaded into OpenClinica. An example spreadsheet is provided in Appendix 3 and a screenshot of an eCRF is shown in Appendix 4. **I was responsible for overhauling how data was captured in OpenClinica after starting my PhD. This included creating open eCRFs (such as the drug and family history eCRFs which required continuous updating) and refining the templates for all the other eCRFs. In addition to correcting errors in the templates, the updates made data entry and analysis more efficient.**

The eCRFs were designed to minimise the use of free text. Drop boxes were provided where possible to allow selection of specific responses. Discrepancy notes could be created to record relevant information not captured in the standard eCRF fields or identify changes to data. **I reviewed the first batch of discrepancy notes entered into OpenClinica and noted that entries into the discrepancy notes were haphazard and sometimes overly complicated. Consequently, I introduced a "controlled vocabulary" and standard operating procedure (SOP) to use when entering discrepancy notes to speed up entry and improve the analysis of the discrepancy notes (Appendix 5).**

A list of the eCRFs is shown below (Table 4) and a list of all the data items captured in OpenClinica is provided in Appendix 6. Most eCRFs captured information from a specific event (e.g. diagnosis, Cohort visit 1, visit 2 …), while others could be updated at any time (e.g. drug therapies, risk factors, family history). The choice of data to be captured in OpenClinica was made collaboratively with physicians from the UK pulmonary hypertension centres.

Table 4. Overview of data collected in OpenClinica for the NIHR BRIDGE PAH Study

Type of eCRF	Name of eCRF	Information captured
Diagnosis	Arterial blood gases	
	Body system	Height and weight
	Clinical blood tests	
	Clinical features at diagnosis	
	Clinical features by examination	E.g. clubbing, Raynaud's
	Demographics	Gender, date of birth and ethnicity
	Echocardiogram	
	Electrocardiogram	
	Exercise performance	Walk test and cardiopulmonary exercise testing
	Functional class	
	Haemodynamics	Diagnostic right heart catheter ± vasodilator challenge
	ID capture	Diagnosis, BRIDGE and Cohort Study IDs
	Imaging investigations	
	Lung function	Spirometry, lung volumes, KCO, overnight oxygen saturation
Continuous	Associated diseases	e.g. presence of congenital heart disease, connective tissue disease
	Cohort relatives	OpenClinica ID of relatives recruited to the Cohort Study
	Drug treatment (Other)	
	Drug treatment (PAH)	
	Family history	
	Risk factors	e.g. history of exposure to anorexigens, recreational drugs, pregnancy
Suspension		Death, transplantation or other reason for suspension including date

Data entry

Research staff at each recruiting centre were responsible for data entry for patients recruited at their centre. Access to OpenClinica was on a named person basis and password protected. Accounts were only provided following a training session on the use of OpenClinica *[study coordinators]*. SOPs for using OpenClinica were created for reference and accessible to all sites. Individuals outside the core study team (i.e. the computational genomics group and study coordinators) were only given access to records of patients from their site.

Diagnostic data for most patients recruited to the NIHR BRIDGE PAH Study was entered retrospectively from source data. Longitudinal follow up data for the Cohort Study was entered prospectively. Local hospital databases, including those storing radiographic images, as well as paper and electronic medical records served as the source data. Diagnostic data was defined as clinical information obtained prior to the commencement of pulmonary artery vasodilator therapies. Longitudinal follow up data for the Cohort Study was defined as clinical information obtained six weeks either side of a study visit date.

Survival data for UK patients were obtained from recruiting centres through the NHS National Spine and local databases. A census of all patients was performed in November 2016. In addition to verifying whether a patient was alive or dead at the time, centres were also asked to provide the date of last contact with patients. To ensure the accuracy of the survival data, if a patient had not been seen for over one year, centres were asked to verify if they had been lost to follow up or had died.

The European recruiting centres, with the exception of the VU University Medical Center, Amsterdam, did not have a named research team for the study. These European centres provided Microsoft Excel spreadsheets containing a limited range of phenotype data. Information in the spreadsheets were processed to ensure it was compatible with the OpenClinica database. This included conversion of values in different units and standardisation of categorical variables. This information was uploaded directly into the OpenClinica database *[Matthias Haimel]*.

I assessed the completeness of the data entered into OpenClinica. This involved analysing the amount of data entered for each data item and assessing any patterns in missing data (e.g. lack of data being entered by a particular site). I looked specifically for missing data in demographic, haemodynamic or functional variables and asked each site to double check if these were available to maximise the amount of data available for analysis.

Data storage

Data entered into OpenClinica was stored on secure servers maintained by Cambridge University School of Clinical Medicine Computing Service and the computational genomics group. Access to the servers was on a named person basis with password protection and only through the Cambridge University Clinical School network. Data was stored in a SQL database and extracted using MySQL (Oracle, USA) *[computational genomics group]*.

R (www.r-project.org) was used to process the extracted data prior to analysis *[computational genomics group and myself]*. I wrote scripts that combined data from OpenClinica with additional data provided in Microsoft Excel spreadsheets (e.g. phenotype data from the European centres), created secondary clinical variables (e.g. calculating age of disease onset and identifying pulmonary artery vasodilator responders based on haemodynamic criteria), and merged subject identifiers (IDs; OpenClinica IDs, BRIDGE IDs and Cohort IDs). The R script I wrote for these functions is provided in Appendix 7**.** The output was a single R data object that was used for further analysis, including data cleaning.

Data cleaning and verification

Several processes were put in place to ensure the accuracy of the phenotype data:

1) Training and supervision:
    a. Training was provided to all research staff regarding data entry *[study coordinators: Carmen Treacy, Katherine Yates, Jennifer Martin]*
    b. SOPs were created and circulated to all sites
    c. Monthly teleconferences between sites were held to facilitate sharing of best practice and highlight any problems *[study coordinators]*
2) Monitoring visits:

a. Each UK centre was visited by the study coordinators to review 10 % of recruited patients. A report was generated to highlight problems and further training given on site if required.

3) **I devised and carried out data error checks**:
    a. The extracted data was assessed for:
        i. Implausible results (Appendix 8; e.g. mPAP > systolic pulmonary artery pressure [sPAP])
        ii. Outliers: defined as 1.5 times the interquartile range (IQR) below the 1st quartile and 1.5 times the IQR above the 3rd quartile
        iii. Anomalous results flagged up by visual inspection of plotted distributions of continuous variables
        iv. Anomalous results flagged by assessing the relationship between pairs of variables (Appendix 8; e.g. visual assessment of plots of FEV_1 and FEV_1 % predicted, or mPAP and sPAP).
        v. Gender differences between self-declared and WGS inferred gender
    b. Errors or queries identified by downstream analyses

Results that were identified as either being incorrect or an outlier were collated into a master spreadsheet. I assigned a unique ID to each query (consisting of a patient's OpenClinica ID, the visit number [e.g. diagnosis, visit 1, visit 2… etc.], the variable name and the initial value entered). This was used to exclude previously verified outliers from future rounds of checks.

The queries were circulated to the appropriate centre to verify each result. Until a result was verified by the centre its was omitted from further downstream analysis. Results highlighted as an outlier may be incorrect or reflect an important clinical characteristic. Outliers may also overly influence the results of regression models. Therefore, if they were identified as influencing the result of such models they were excluded, and the models rerun (further details below).

Despite these efforts the possibility of finding errors in the data during further analysis remained. Ad hoc queries, arising from downstream analyses, were handled in the same way as the outliers and errors described above.

Computed tomography imaging of the lungs

CT images of the chest taken at the time of diagnosis (in a subset of patients that are defined below) were obtained from each centre where available. The images were sent to Papworth Hospital and the Royal Hallamshire Hospital after being anonymised and burnt to compact disc. The images were reported centrally by two expert radiologists in pulmonary hypertension *[Dr Nicholas Screaton and Dr Andrew Swift]*, who were blinded to the diagnosis and genotype. The CT images were assessed based on a customised prospectively designed proforma (Appendix 9). The images were read independently and a consensus diagnosis (PAH or any suspicion of PVOD / PCH) reached for each patient.

CT images for patients with biallelic *EIF2AK4* variants from all centres were assessed and compared to CT images from patients with *BMPR2* variants or with no variants in previously identified disease associated genes from Papworth Hospital and the Royal Hallamshire Hospital. Only patients with normal spirometry (FEV_1 and FVC > 80 %, FEV_1 / FVC > 0.7) were selected for the latter two groups.

Histology

The explanted lung tissue of one patient with a clinical diagnosis of idiopathic PAH and biallelic *EIF2AK4* variants was available for further analysis. Four micrometre (µm) tissue sections were cut from formalin-fixed paraffin wax embedded blocks from the explanted lung tissue *[Dr Mark Southwood]*. Representative sections from each lobe of both lungs were stained with Haemotoxylin and Eosin and Elastic-Van Gieson stains. Two expert histopathologists examined the sections independently by light microscopy *[Dr Peter Dorfmüller and Dr Stephen Preston]*.

Whole genome sequencing

Sample preparation

Whole blood samples, from patients recruited to the NIHR BRIDGE Study cohorts, were sent to the Cambridge Translational Genomics (CATGO) laboratory for central processing. Two six millilitre whole blood samples collected in ethylenediaminetetraacetic acid (EDTA) tubes were required for each patient. Genomic DNA was extracted from the samples then their

concentration and quality assessed by Qubit and gel electrophoresis respectively. The samples were sent to Illumina for preparation of DNA libraries (PCR free). The libraries underwent next generation sequencing using 100 to 150 bp paired-end sequencing on Illumina HiSeq 2500 and HiSeq X platforms (Illumina Inc, San Diego, USA).

Bioinformatics pipeline

The processing and annotation of the WGS data was performed by the NIHR BRIDGE Study bioinformaticians. Their contributions are specified in brackets below. Batches of FASTQ files were returned by Illumina to the BRIDGE Study core bioinformatics team for further processing into merged VCF files. Reads were aligned against GRCh37 and variants were called using the Issac Aligner and Variant Caller respectively (version 2, Illumina Inc.). Genebuilds for genes previously associated with PAH (*BMPR2*, *ACVRL1*, *ENG*, *SMAD1*, *SMAD4*, *SMAD9*, *TBX4*, *KCNK3*, *CAV1*, and *EIF2AK4*) were based on Ensembl v75. Variants in these genes were extracted and annotated using Ensembl's Variant Effect Predictor (VEP) v84 *[Dr Stefan Gräf]* (335). VEP was also used to annotate data from the Exome Aggregation Consortium's (ExAC) database (330).

Deletions (resulting in the loss of more than 50 bp) were identified by applying Isaac copy number variant caller (Canvas, Illumina) and Isaac Structural Variant Caller (Manta, Illumina) *[Matthias Haimel]*. To be called by both Canvas and Manta deletions required a reciprocal overlap of ≥ 20 %. Overlapping deletions represented in the Zarrei dataset with a reciprocal overlap of ≥ 50 % and deletions with a non-PAH BRIDGE control frequency of more than 1 in 1,000 were excluded (336).

WGS inferred ethnicity assessment

The Genesis package in R was used by Marta Bleda to estimate the ethnicities of patients recruited to the study. Initially, 2100 patients recruited to the 1000 Genomes Project with known ethnicity (European, African, East-Asian, South-Asian) were assessed based on a selection of 35,114 autosomal SNPs that were present in both the 1000 Genomes Project dataset and the NIHR BRIDGE Study, were fully sequenced in the BRIDGE dataset, and had a minor allele frequency of more than 0.3. PCA was performed on these SNPs and the loadings for the five leading principal components taken forward. The genotypes of the NIHR BRIDGE

PAH patients were compared to this and the likelihood of the individuals recruited to the study belonging to each of the 4 ethnicities calculated. The ethnicity with the highest likelihood was used as the WGS inferred ethnicity for that individual.

Identification of unrelated individuals

The PRIMUS software package was used to identify non-related individuals (i.e. no closer than 3rd degree relationships) amongst both non-PAH BRIDGE control subjects and PAH patients *[computational genomics group]* (337). The number of unrelated non-PAH control subjects from other BRIDGE Study cohorts was maximised by including either patients with other rare diseases or their unaffected relatives.

Variant filtering

Likely causal variants were identified based on minor allele frequency (MAF) and predicted deleteriousness. Variants were considered further if they had a MAF of less than 1 in 10,000 in unrelated non-PAH BRIDGE controls, the ExAC database and the UK 10K Study (330, 331). Rare variants that passed the MAF filtering step were then assessed for deleteriousness. Variants were considered pathogenic based on a CADD score $\geq$ 15 and PolyPhen-2 or SIFT predictions not classified as “benign” or “tolerated” respectively (332-334).

Statistical analysis

I performed all the statistical analyses in R. R was also used to create the graphs presented in the Thesis. Annotated examples of the R code I used in the statistical analyses are provided in Appendix 10.

Summary statistics

Summary data is presented as median and IQR unless otherwise stated. Percentages were calculated using the number of subjects with recorded observations as the denominator.

Group comparisons

Differences between groups for categorical variables were assessed using Fisher’s exact test. Fisher’s exact test was selected (over the Chi square test) given that some features may be

rarely represented in the dataset (e.g. hyperlipidaemia in young patients). The Cochran-Armitage test was applied, using the chisq_test function from the "coin" package, to assess between group differences for ordinal (ordered categorical) variables (338). Continuous variables were assessed for normality by performing the Shapiro-Wilk test and visually assessing histograms of their distributions. Non-parametric variables were assessed using the Mann-Whitney U / Wilcoxon rank-sum test (2 comparator groups) and the Kruskal-Wallis test (3 or more comparator groups). While parametric variables were assessed using the Student t test or ANOVA. Post-hoc pairwise comparisons were performed using Dunn's test with Bonferroni corrections applied.

Regression models

Simple linear models were created to assess associations between variables. Dependent variables were transformed to normalise their distribution where required (square and square root transformations). Independent variables that were associated with the dependent variable with a $p < 0.1$ in simple models were used to create multiple linear models. Multiple collinearity between independent variables was assessed by calculating the variance inflation factors and assessing correlation matrices. Residuals plots for the models were visually inspected to confirm the suitability of the models and identify outliers with an extreme effect on the model.

If the dependent variable could not be transformed to fit a normal distribution rank regression models were used to test for associations between variables. Rank regression uses the ranking of the data point in the dataset rather than its actual value for analysis. Therefore, they can be used to analyses variables that are not normally distributed. They are also less affected than linear models by extreme outliers. Multivariate rank regression models were created to assess differences between groups using the Rfit package (339). The Bent3 score was used to optimise the models for the skewed distributions. Models were checked by visually assessing q-q plots of the studentised residuals.

Where missing data resulted in the loss of large numbers of subjects in multiple regression models these variables were excluded from the models.

Survival analyses

Semi-parametric Cox-proportional hazards models were used to assess survival using the "survival" package in R (340). Survival was assessed from both diagnosis to death (in which patients were right censored at transplantation) and diagnosis to death or transplantation. Covariates such as age at diagnosis, gender and genotype were used in the models.

The proportional hazards assumptions were tested by assessing Schoenfeld residuals over log time (341). The goodness of fit of the model was assessed by plotting the log of cumulative hazard of the Cox-Snell residuals against the log of time and confirming the simple regression has a 0 intercept and a slope of 1 (342).

The inclusion of retrospectively recruited and prevalent patients in a survival analysis assessing time from diagnosis to death / transplantation can cause immortal time bias. The immortal time is the period between diagnosis and enrolment in the study. Prospectively recruited patients must have survived up until recruitment. However, patients with a worse prognosis diagnosed at a similar time point may not have survived long enough to enrol in the study. The inclusion of retrospectively recruited patients introduces a further selection bias. To understand the effect of these biases, left truncated survival analyses were performed on UK patients recruited prospectively to the study. In these analyses, the survival period was defined as the time from diagnosis to death and patients only entered the risk set after enrolment into the study (consent date). This applied to all patients included in the left truncated analyses, so no selection bias was introduced and nor was the immortal period differentially misclassified.

Over representation analyses

Differences in the frequency of *EIF2AK4* variants in a heterozygous state between PAH patients and non-PAH control subjects was assessed using the Fisher's exact test. For this analysis only variants in unrelated subjects were included. The frequency of rare and predicted deleterious *EIF2AK4* variants in a heterozygous state in PAH index cases was also compared to publicly available information in the ExAC database (http://exac.broadinstitute.org) (330). This analysis provides the maximum estimate of the

frequency of *EIF2AK4* variants in a heterozygous state in the ExAC database as variants in ExAC were assumed not to be in a compound heterozygous state.

Results 1: Description of the NIHR BRIDGE PAH Study

Introduction

PAH is a heterogeneous group of related conditions defined haemodynamically by a mPAP ≥ 25 mmHg and a PCWP ≤ 15 mmHg (1). PAH forms group 1 of the current clinical PH classification (Table 1) and is sub-classified based on disease pathogenesis (i.e. idiopathic, heritable, drug / toxin induced and associated forms). The classification is of clinical relevance, despite treatment options being similar for each subgroup, as the different subgroups have significantly different prognoses. For example, patients with heritable PAH and variants in *BMPR2* and PAH associated with connective tissue diseases have a significantly worse prognosis compared to patients with idiopathic PAH (30, 343). In contrast, patients with exposure to anorexigens have a similar prognosis compared to those with idiopathic PAH (162, 344).

Large national and international collaborations have set up registries of patients with PH to assess their clinical characteristics and long-term outcomes (25, 183, 192, 202, 266, 269, 279). The European Union Guidelines on Pharmacovigilance for Medicinal Products for Human Use define a registry as "a list of patients presenting with the same characteristic(s)" (345). This definition is of importance as these registries have differing inclusion and exclusion criteria. Differences are found in the diagnoses included, the haemodynamic criteria for inclusion and/or the inclusion of prevalent as well as incident cases. For example, the REVEAL study includes all patients with PAH, whereas the French registry limits its inclusion criteria to idiopathic, heritable and anorexigen associated PAH (183, 266, 279). Although this may lead to difficulties in direct comparisons it also provides the ability to contrast different populations of patients. The characteristics of some of the major registries are summarised in Table 3.

Registries provide "real world" information regarding the assessment and follow up of patients with PAH. There is heterogeneity in assessments even within individual registries. Such heterogeneity may arise from subjectivity on the part of clinicians or from differences in guidelines and assays used in hospitals. As an example, significant variation in functional class determined by clinicians has been reported (346). Variability is also present in more objective

assessments, such as the measurement of PCWP, despite the availability of consensus guidelines (347). In part, this variation in practice increases the generalisability of registry derived data in comparison to data from randomised control trials. However, such variation also needs to be assessed as a potential confounder. A further drawback of registry data is that due to their observational nature datasets for individual patients may be incomplete.

The inclusion of prevalent patients is a contentious issue in registries especially when assessing survival. The inclusion of such patients can lead to survival bias (whereby prevalent patients are likely to have less severe disease and therefore a better prognosis) and immortal time bias (whereby there is a period of time prior to the start of the study that is unobserved and patients with a poor prognosis would not have survived to be recruited in to the study). Yet, in a rare disease such as PAH, the recruitment of incident patients alone necessitates a prolonged recruitment period and a longer follow up period in order to observe enough events to sufficiently power any survival analysis. Alternatively, statistical methods, such as left truncation that delays the entry of prevalent patients into the risk set, can minimize these biases (348).

PAH registries have sought to assess characteristics of prognostic significance (Table 3). However, most variables identified to be of prognostic significance have not been validated across all studies. The heterogeneity in study design and insufficient study power may in part explain this lack of concordance. Variables shown to be of prognostic significance across several studies include age, gender, functional class, 6mwt distance, NT-ProBNP, pericardial effusions, D_{LCO}, RAP, CO, CI and PVR (182, 183, 269-272). Risk stratification calculators have been formulated to identify patients with poor outcomes using these prognostic variables (1, 266, 349, 350). Early identification of patients with poor outcomes can lead to more aggressive upfront treatment and/or early referral for lung transplantation assessment. Although, many registries have sought to identify variables of prognostic significance, only a few studies have sought to describe idiopathic PAH patients by dichotomising them based on these variables (351).

The BRIDGE PAH Study is a prospective study enrolling patients with PAH to elucidate the genetic basis of the disease through WGS and phenotyping. The study shares many of the features of a registry described above. In this chapter I will aim to:

1) Describe the quality of the phenotype data captured as part of the study.
2) Describe the demographics and disease aetiology of patients recruited to the study.
3) Identify characteristics of prognostic importance in idiopathic PAH.

Phenotype data verification

Before detailed analysis of the diagnostic phenotype data collected as part of the BRIDGE PAH Study, I assessed the data for accuracy and completeness. A data freeze was performed on 14/10/2017. Only patients with WGS data available at this time were taken forward for further assessment and phenotype data captured in OpenClinica by this date was used.

One-thousand one-hundred and thirty-one patients recruited to BRIDGE PAH Study had been sequenced by the time of the data lock. On review, 61 patients were determined not to meet the inclusion criteria for the study and were excluded. For the remaining 1,070 patients, 108,218 data points had been captured in OpenClinica. On average this equated to 101 data points per patient recruited to the study. Given the retrospective nature of data collection, manual data entry and variation in practices between centres I systematically assessed data entry to maximise data capture and ensure its accuracy for downstream analyses.

Data completion

The amount of diagnostic data entered into OpenClinica varied between the recruiting centres (Table 5). In particular, the European centres that did not have dedicated research staff for the study provided limited phenotypic data. The exception to this was the VU Medical Center in Amsterdam that had named research nurses responsible for data entry directly into OpenClinica.

Great Ormond Street Hospital, a paediatric specialist centre, entered less data compared to the UK adult pulmonary hypertension centres. This reflects differences in the assessment of paediatric patients, for example exercise testing and lung function testing are not routinely performed (Figure 2). The data completion rates for all diagnostic data from all recruiting centres are provided in Appendix 11 and summarised in Figure 3.

Table 5. Amount of diagnostic data collected in OpenClinica

	Number of patients recruited	Number of data points captured in OpenClinica	Average number of data points captured per patient
Royal United Hospital, Bath	23	3363	146.2
Royal Free Hospital, London	29	4117	142.0
Hammersmith Hospital, London	249	34224	137.4
Royal Jubilee Hospital, Glasgow	80	10534	131.7
Royal Brompton Hospital, London	59	7452	126.3
VU University Medical Center, Amsterdam	45	5602	124.5
Freeman Hospital, Newcastle	59	7331	124.3
Royal Hallamshire Hospital, Sheffield	118	13898	117.8
Papworth Hospital, Cambridge	120	13910	115.9
Great Ormond Street Hospital, London	10	1009	100.9
Slovenia*	3	253	84.3
University of South Paris	144	4541	31.5
San Matteo Hospital, Pavia	49	989	20.2
University Hospital Giessen	77	995	12.9

*3 paediatric onset patients

Centre	Arterial blood gas	Body system	Blood tests	Examination	History	Echo	ECG	Exercise test	Right heart catheterisation	Vasodilator study	Imaging	Lung function
Golden Jubillee Hospital	4.1	4.4	36.1	10.6	6.0	7.2	9.0	10.7	17.8	5.9	8.4	9.5
Great Ormond Street Hospital	0.5	4.4	30.3	10.3	6.1	8.7	7.4	2.2	16.3	6.6	3.4	1.3
Hammersmith Hospital	3.2	3.4	45.2	11.3	6.3	7.7	8.9	12.1	14.9	3.4	8.9	11.2
Freeman Hospital	3.7	4.8	35.3	9.9	6.9	6.5	7.3	10.8	17.4	4.2	3.4	9.0
Papworth Hospital	1.7	4.9	34.9	10.3	6.6	6.9	5.4	7.7	16.2	2.1	5.7	8.6
Royal Brompton Hospital	2.9	4.6	38.6	10.3	6.5	7.2	7.1	8.7	14.0	7.2	8.8	7.4
Royal Free Hospital	1.7	5.0	44.7	10.3	6.9	7.0	7.5	8.1	19.2	9.4	8.1	10.0
Royal Hallamshire Hospital	0.9	4.7	32.2	10.4	6.8	5.1	7.6	6.6	17.0	7.0	9.7	5.7
Royal United Hospital Bath	0.8	5.0	47.9	11.6	7.0	10.0	10.3	9.6	20.6	4.3	10.2	10.4
San Matteo Hospital	0.0	0.0	0.0	0.0	1.0	0.0	0.0	1.6	6.7	0.9	0.0	0.0
Slovenia	0.0	5.0	8.3	11.0	6.7	6.7	7.0	4.3	15.0	8.7	4.3	0.7
University Hospital Giessen	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	2.9	0.0	0.0	0.0
University of South Paris	0.0	0.0	0.0	0.0	5.8	1.0	0.0	3.5	8.3	0.0	0.0	0.0
VU University Medical Center Amsterdam	2.2	4.8	30.2	8.5	6.6	5.8	7.7	16.1	16.1	8.1	9.7	5.6

Colour													
Centre Rank	1st - most data	2	3	4	5	6	7	8	9	10	11	12	13	14th - least data

Figure 2. Differences in data collection by recruiting centre

The amount of data collected in OpenClinica was assessed. The mean number of data items captured and entered into OpenClinica for each eCRF for each centre was calculated. This value is presented in the figure. Each column is coloured according to rank with the centre collecting the least data for a particular eCRF coloured red and the centre collecting the most data coloured blue. The European recruiting centres provided much less phenotypic data than the UK centres.

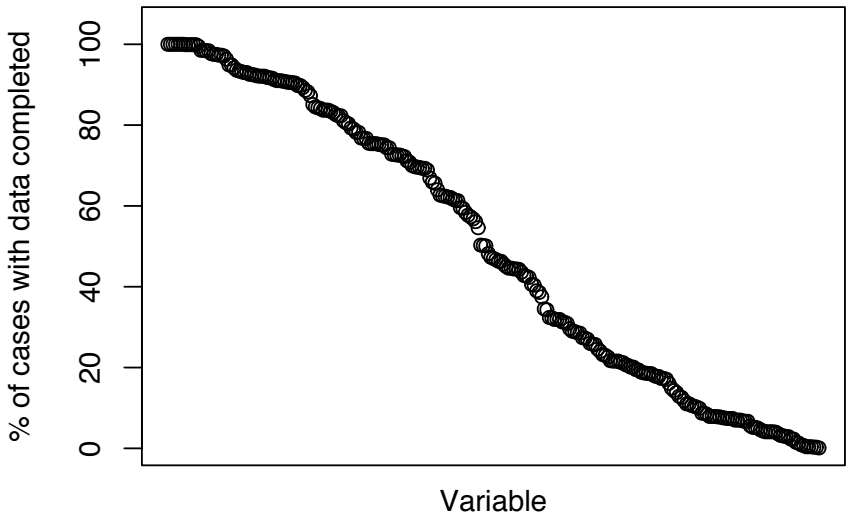


Figure 3. Summary of the completeness of data collection in OpenClinica

The amount of data captured in OpenClinica was assessed for each data item (or variable; n = 257). The figure is generated from the data presented in Appendix 11, which provides the name and amount of data collected for each item. Approximately one-third of the data items requested in OpenClinica have a data completion rate of more than 75 %. A similar proportion of data items have a data collection rate of less than 25 %. Future studies may choose to exclude poorly reported data that are less informative and focus on better reported data items.

Given the discrepancy between UK and European centres, only data completion for the UK pulmonary hypertension centres was assessed further. The data completion rates for selected data items (variables) are shown in Table 6. Data items related to cardiopulmonary exercise testing, acute pulmonary artery vasodilator challenge testing and z-scores of pulmonary function tests were the least well reported. Only 33 % of data items had completion rates more than 75 %. A similar proportion (34 %) had completion rates of less than 25 %.

Even amongst UK centres there were differences in the amount of data entered. This may represent variation in clinical practice and record keeping between centres. As an example, echocardiography is no longer routinely used to assess patients referred to the Royal Hallamshire Hospital, Sheffield as this has been replaced with cardiac MRI. Missing data in what were considered core variables (gender, date of birth, date of diagnosis, mPAP, PCWP, CO, functional class and 6mwt distance) were queried with the centres. Despite this, PCWP, a haemodynamic criterion required for the diagnosis of PAH, was only available for 81 % of patients. This may reflect difficulties in obtaining an accurate PCWP (352). In order to

maximise the number of patients included in the study a missing PCWP was accepted if the expert centre felt that the clinical diagnosis was consistent with the inclusion / exclusion criteria.

Table 6. Data completion rates for selected variables from UK centres		
Item name	**Number complete**	**% complete**
Date of birth	747	100.0
Centre	747	100.0
Gender	747	100.0
Diagnosis	746	99.9
Ethnic category	746	99.9
Weight (kg)	709	94.9
Date of right heart catheterisation	709	94.9
Functional class	704	94.2
mPAP (mmHg)	698	93.4
WBC ($x10^9$/l)	695	93.0
Hb (g/l)	691	92.5
Platelets ($x10^9$/l)	690	92.4
Height (cm)	689	92.2
sPAP (mmHg)	688	92.1
dPAP (mmHg)	686	91.8
Creatinine (μmol/l)	681	91.2
Systemic diastolic BP (mmHg)	677	90.6
Systemic systolic BP (mmHg)	677	90.6
CO (L/min)	675	90.4
RAP (mmHg)	667	89.3
Resting S_aO_2 (%)	652	87.3
6mwt distance (m)	636	85.1
PVR (WU)	625	83.7
FEV_1 (L)	625	83.7
FVC (L)	621	83.1
S_vO_2 (%)	617	82.6
FEV_1 (% predicted)	615	82.3
PCWP (mmHg)	606	81.1
FVC (% predicted)	600	80.3

KCO (mmol/min/kPa/l)	561	75.1
KCO (% predicted)	532	71.2
NT-ProBNP (ng/l)	194	26.0

6mwt – six minute walk test, BP – blood pressure, CO – cardiac output, dPAP – diastolic pulmonary artery pressure, FEV_1 – forced expiratory volume in 1 second, FVC – forced vital capacity, Hb – haemoglobin, KCO – transfer coefficient for carbon monoxide, mPAP – mean pulmonary artery pressure, NT-ProBNP – N terminal Pro brain natriuretic peptide, PCWP – pulmonary capillary wedge pressure, PVR – pulmonary vascular resistance, RAP – right atrial pressure, S_aO_2 – peripheral arterial oxygen saturation, sPAP – systolic pulmonary artery pressure, S_vO_2 – mixed venous oxygen saturation, WBC – white blood cells

Monitoring visits

The BRIDGE PAH / Cohort Study coordinators have monitored each UK centre twice since the study began. I analysed the monitoring report forms generated to assess the errors picked up during the monitoring visits. The monitoring reports documented errors identified in OpenClinica but did not always specify which data fields were verified as correct. The stated aim was to check all diagnostic data fields for patients selected for monitoring. Therefore, it was not possible to quantify accurately which variables in OpenClinica were checked. To improve this and allow better quantification of errors, the discrepancy note function in OpenClinica will be used to document which variables are checked in future monitoring visits.

During the monitoring visits, 85 patient records (8 % of all patients recruited and 11 % of UK patients) entered in OpenClinica were compared to source documentation. In total 74 errors were identified, all of which were amended. Thirty errors related to right heart catheterisation data. The Royal Brompton Hospital had 4 queries related to the clinical diagnosis and 2 of these patients were subsequently excluded from the study. The identification of these errors at the Royal Brompton Hospital resulted in changes to the oversight of the research nurses and review of all patients recruited to the study by both the research nurses and the principle investigator.

Missing data for the core variables (defined above) were identified in 15 patients and amended. Three-hundred and seven eCRFs, from 75 patients, were identified as having accurate records.

Data verification

To further improve the accuracy of the phenotype data entered into OpenClinica I checked the data for errors and outliers (Appendix 8). Items flagged up by these checks were sent to the centres for verification and correction in OpenClinica if required. In total 2,017 queries were raised. This consisted of 85 errors, 90 unit discrepancies, 5 gender discrepancies (WGS inferred gender different from the gender entered into OpenClinica), 24 date queries and 1,813 outliers. A total of 1,486 queries (74 %) were checked by the centres, 437 queries (22 %) were still pending verification and 94 queries (5 %) were not checked as the source

documentation was no longer available. From the queries that were checked by the centres 262 (18 %) resulted in a change to the data entered in OpenClinica. Data related to queries that could not be confirmed or were still pending confirmation were removed from further analysis.

After I had verified these queries with the recruiting centres a number of critical discrepancies remained. There was a mismatch between WGS inferred gender and the gender entered into OpenClinica for 3 patients. There was no history of gender reassignment in these patients. It is likely that these DNA samples were incorrectly assigned. Consequently, these patients were excluded from further analysis.

Nine patients who had a mPAP < 25 mmHg were also excluded from further analysis. Forty-two patients had a PCWP and/or a left ventricular end diastolic pressure (LVEDP) > 15 mmHg, these patients were excluded from subsequent analyses. Another 213 patients did not have a recorded PCWP or LVEDP but were included in the study as they were confirmed to have a clinical diagnosis consistent with the inclusion / exclusion criteria. A further 66 patients did not have a mPAP recorded but were also included in the analysis as they were confirmed to have a clinical diagnosis consistent with the inclusion / exclusion criteria.

Differences between recruiting centres

An important confounder to take into account was variation in practices, protocols and assays between the different recruiting hospitals. Given the limited data that was provided by most European centres, I assessed for differences amongst the recruiting UK centres and the VU Medical Center, Amsterdam. After correction for multiple testing 55 variables remained significantly different between centres. These are summarised in Appendix 12 and Figure 4.

	Age	RAP	sPAP	dPAP	mPAP	PCWP	CO	CI	SvO2	RV diastolic volume	RVEF	Abnormal RV on echo	TAPSE
Glasgow	52.3	8	87	32	52	8	3.8	2.1	62.9	191	39	70	1.4
Great Ormond Street	7.8	5	46.5	16.5	29.5	7.5	3.1	3	69.3	98.5	58	8	1.9
Hammersmith	47.5	10	85	33	54	10	3.7	2	64	203.5	35	218	1.5
Newcastle	59.1	8	81	28	48	8	4	2.1	65	223	34.5	47	1.6
Papworth	47.8	8	88	35	53	10	3.5	2	66.5	201	36	96	1.6
Royal Brompton	42.9	6	79	35	53	7.5	3.9	2	64.7	190	45	46	1.6
Royal Free	46.7	9	91	35	54	11	4.2	2.4	66	197	33	21	2
Bath	62	11	91	38	57	14	3.9	2	61.9	119.5	46	21	1.6
Sheffield	51.3	10	87	36	55	10	4	2.2	62.4	146.7	37	74	1.4
Amsterdam	45.3	7.5	78	33	50	8	4.6	2.4	67.5	142	40	18	1.5

	Syncope	RBBB	Dominant R wave	Vasoresponder	6mwt distance	FVC (% predicted)	FEV1 (% predicted)	Resting SaO2	Pre walk SaO2	Post walk SaO2	Height	Weight	BMI	BSA	Systemic SBP	Systemic DBP
Glasgow	28	15	30	7	320	97	87	94.5	94.5	87	163	78.4	28.7	1.9	121	74
Great Ormond Street	7	0	1	0	240	56	52	98	95	97	121.2	21.6	16.2	0.8	93.5	57.5
Hammersmith	65	53	86	2	280	89	82.1	95	96	93	165	73	27.6	1.9	126	78.5
Newcastle	6	7	15	2	313	92	80.5	94	94	89	162.5	75	28.3	1.8	118	72
Papworth	27	10	11	7	343	100	88	96	96	91	163	74.1	27.1	1.8	120	76
Royal Brompton	15	9	7	0	315	93.2	81.2	95	97	92	164	77.5	26.8	1.8	117	74
Royal Free	7	4	10	1	352	98.3	80.3	95	96.5	94.5	165	73	27.1	1.8	118	78
Bath	5	2	13	0	209	101	90	95	96	92	163	87	31.6	1.9	125	82
Sheffield	24	11	15	11	320	98.3	87.5	96	95	86	165	78.5	27.9	1.9	124.5	78.5
Amsterdam	5	2	20	2	475	102	96	97	97	93	169	72	24.3	1.8	114	80

Colour										
Centre rank	1st - highest value	2	3	4	5	6	7	8	9	10th - lowest value

Figure 4. Significant difference between patients under the care of different recruiting centres

Differences in the characteristics of patients from different recruiting centres were assessed. Fisher's exact test was used to assess differences in categorical variables and the Kruskal Wallis test was used to assess differences in continuous variables. Correction for multiple testing was made with the false discovery rate adjustment. Variables that remained significantly different are summarised in the figure. Each column is independently coloured according to the rank of each centre to aid visualisation of which centres may be responsible for the significant differences. As expected paediatric patients from Great Ormond Street Hospital were responsible for some of the extreme values compared to other centres.

Values indicate the median value (continuous variables) or the percentage of the total (categorical variables) for each recruiting centre.

6mwt – six-minute walk test, BMI – body mass index, BSA – body surface area, CI – cardiac index, CO – cardiac output, DBP – diastolic blood pressure, dPAP – diastolic pulmonary artery pressure, FEV_1 – forced expiratory volume in 1 second, FVC – forced vital capacity, RAP – right atrial pressure, mPAP – mean pulmonary artery pressure, PCWP – pulmonary capillary wedge pressure, RBBB – right bundle branch block, RV – right ventricle,

RVEF – right ventricular ejection fraction, SBP – systolic blood pressure, sPAP – systolic pulmonary artery pressure, TAPSE – tricuspid annular plane systolic excursion.

Figure 4 demonstrates qualitatively that many of the differences identified are due to patients recruited from Great Ormond Street Hospital and Royal United Hospitals Bath (Bath). To quantify which pairwise comparisons were significantly different Dunn tests were performed with Bonferroni corrections. As expected age at diagnosis was significantly different between centres, in particular patients had a significantly younger age at diagnosis at Great Ormond Street Hospital (median [IQR]: 8 years [5 - 10]) compared to all the other centres. Additionally, the age at diagnosis of patients recruited in Amsterdam (45 years [29 - 51]) was younger compared to Glasgow (52 years [41 - 66]), Newcastle (59 years [41 - 69]) and Bath (62 years [48 - 72]; all $p < 0.05$). Brompton (43 years [33 - 54]) also had a younger age at diagnosis compared to Newcastle and Bath ($p < 0.05$). The centres based in London (Great Ormond Street, Royal Brompton, Royal Free and Hammersmith Hospital) and Amsterdam had a higher proportion of non-European patients compared to the other centres. The proportion of female patients was not significantly different between centres.

With regard to pulmonary haemodynamic variables, mPAP was significantly lower amongst patients from Great Ormond Street (30 mmHg [28 - 33]) compared to patients from all other centres except for Newcastle (48 mmHg [39 - 56]). Whereas the PCWP was significantly lower in patients from Glasgow (8 mmHg [5 - 10]) compared to Hammersmith (10 mmHg [8 - 13]), Royal Free (11 mmHg [9 - 12]), Bath (14 mmHg [11 - 15]) and Sheffield (10 mmHg [8 - 13]; all $p < 0.05$). While Bath (14 mmHg [11 - 15]) had higher PCWPs compared to Papworth (10 mmHg [8 - 11]), Hammersmith, Newcastle (8 mmHg [6 - 10]), Brompton (8 mmHg [6 - 11]) and Amsterdam (8 mmHg [7 - 10]; all $p < 0.05$). CI was similar between most centres. Although, patients from Amsterdam (2.4 L/min/m^2 [2.0 - 3.2]) had a higher CI compared to patients from Hammersmith (2.0 L/min/m^2 [1.6 - 2.6]) and Papworth (2.0 L/min/m^2 [1.6 - 2.4]; all $p < 0.05$).

Differences in the demographics of each centre's catchment area and variation in local assays and tests are likely to explain the disparities between the centres demonstrated above. Therefore, in regression models presented in this Thesis, recruiting centre was always

assessed as an independent variable in univariate models. If significant in univariate models recruiting centre was also included as a covariate in multivariate models.

Patient recruitment

A total of 1,327 patients were recruited to the BRIDGE PAH Study; the recruitment target was 1,250. WGS data was available for 1,131 patients (85 %) at the time of the data lock. Sequencing data for the remainder was pending delivery by Illumina. One-hundred and twelve patients were excluded following the data verification and cleaning processes described above. The remaining 1,019 patients were retained for analysis. Of these 1,019 patients, 1,009 were classified as index cases and 10 as affected relatives. A breakdown of the study population is provided in Figure 5.

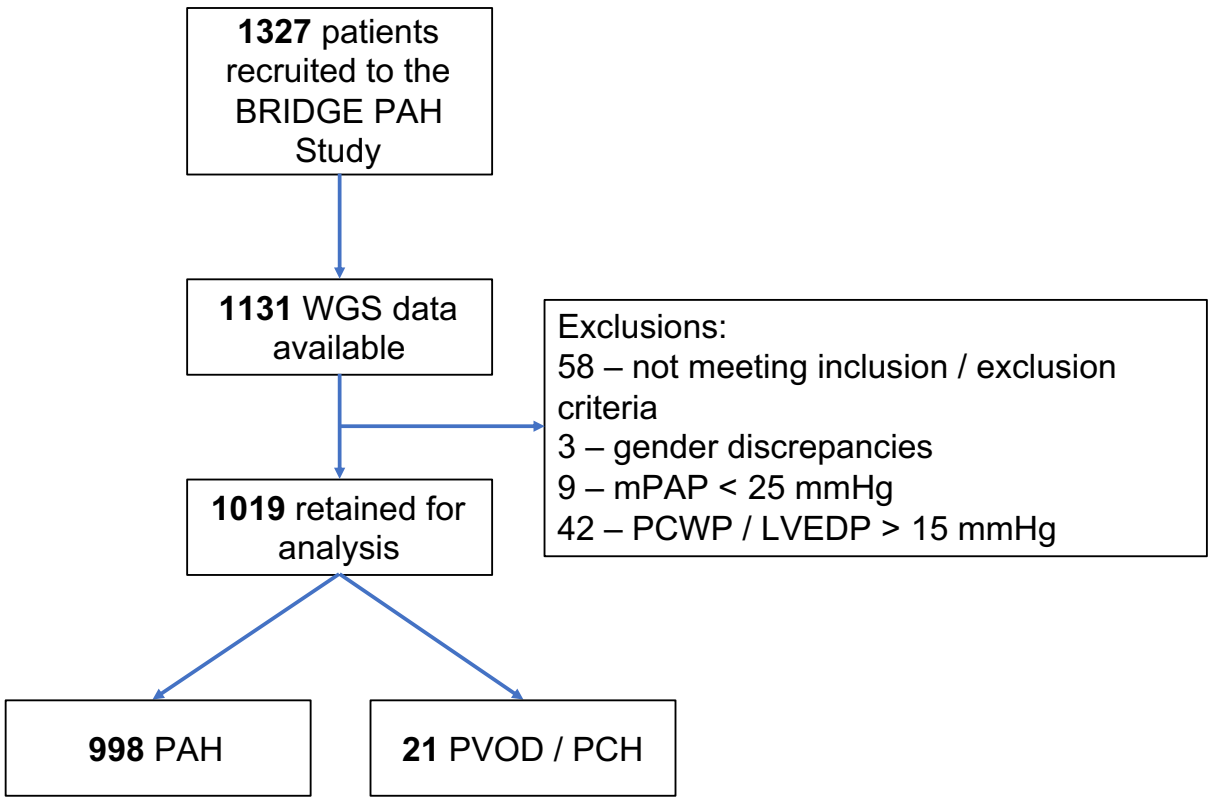


Figure 5. Composition of patients recruited to the NIHR BRIDGE PAH Study and used for analysis

In the results chapters of this Thesis patients are classified as having either familial PAH or idiopathic PAH, based on the presence or absence of a family history of the disease. The term heritable PAH is used when referring to patients with a causal variant in a disease associated

gene *or* those with a family history of the disease. Additionally, the term PAH is used to refer to patients with either heritable, idiopathic PAH or drug/toxin associated PAH. It does not include the other forms of PAH listed in group 1 of the current classification (1).

Demographics

Gender

Sixty-eight percent of the study population were female, resulting in a 2.2 : 1 female gender bias. This bias was not significantly different between patients diagnosed with PAH (2.2 : 1) and PVOD / PCH (1.6 : 1; Fisher's exact test p = 0.489).

Age at diagnosis

The median age of the study population was 49 years [36 - 63] (median [IQR]). This was not significantly different between patients diagnosed with PAH (49 years [35 - 63]) and PVOD / PCH (54 years [42 - 68); Wilcoxon rank-sum test p = 0.172). The age at diagnosis was significantly different between male and female patients. Male patients were diagnosed later in life (53 years [38 - 67]) compared to female patients (48 years [34 - 61]; Wilcoxon rank-sum test p = 0.003). This appears not to be due to a delay in diagnosis in male patients as the time from symptom onset to diagnosis was similar between male (1.2 years [0.6 - 2.5]) and female patients (1.1 years [0.6 - 2.1]; Wilcoxon rank-sum test p = 0.928).

The density plots of age at diagnosis show bimodal distributions for both male and female patients, but this was most obvious for male patients (Figure 6). For female patients, the modes were at 33 years and 49 years. Whereas for male patients, the modes were at 43 years and 68 years. The nadir for male patients, 53 years, was selected to dichotomise the study population into "younger" and "older" groups. Based on this dichotomisation 57 % of patients were classified in the younger group and 43 % in the older group. In the older group, there was still a female gender bias (1.7 : 1) but this was significantly less than in the younger group (2.6 : 1; Fisher's exact test p = 0.003).

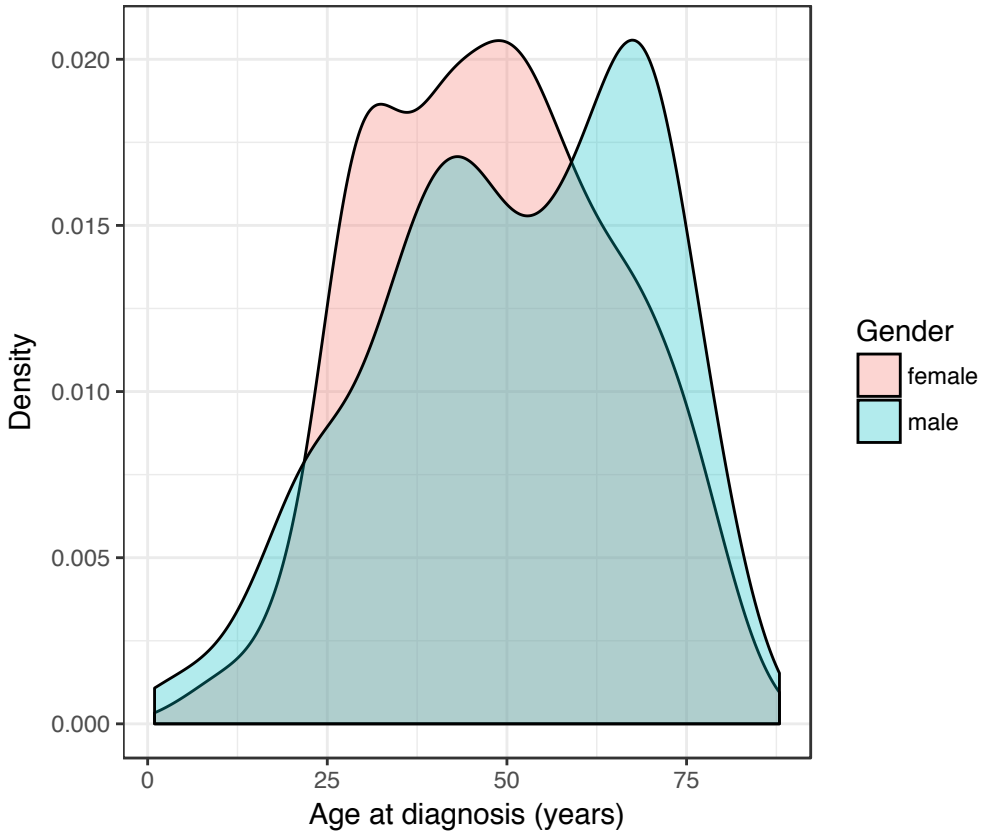


Figure 6. Density plot showing the distribution of age at diagnosis by gender

Age at diagnosis was calculated for all patients (n = 1019).The density plot is chosen to represent the data over a histogram because of the large overlap between the two groups. As age at diagnosis is a continuous variable and there are a large number of observations a density plot is a valid visualisation of the data. A bimodal distribution for age at diagnosis can be seen for both genders. Additionally a higher proportion of male patients (35 %) are diagnosed at an older age (defined as over 53 years) compared to female patients (22 %).

A bimodal distribution for age at diagnosis was seen amongst patients with a diagnosis of PVOD / PCH (Figure 7). These patients have modes at 46 years and 67 years. In contrast, patients with a diagnosis of PAH have a unimodal distribution for age at diagnosis (mode: 45 years) when not dichotomised by gender.

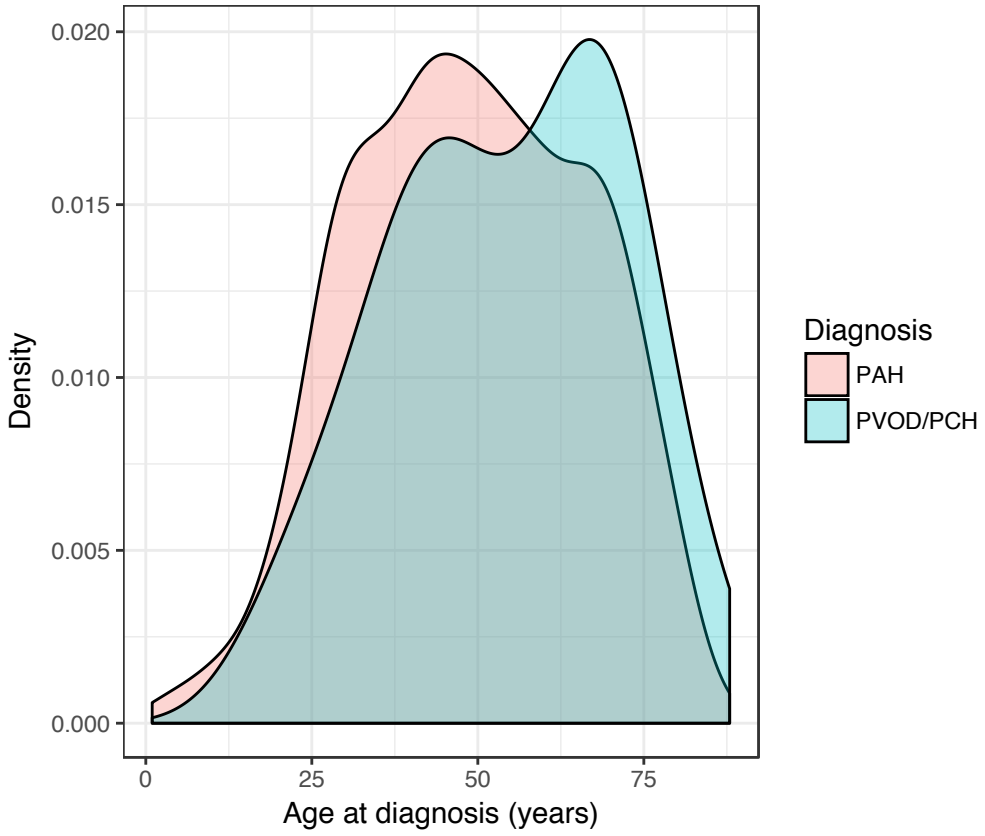


Figure 7. Density plot of age at diagnosis by clinical diagnosis

A bimodal distribution can be seen for the age at diagnosis for patients with PVOD / PCH (modes at 46 and 67 years), perhaps indicative of the distinction between younger patients with a genetic cause and older patients with cardiovascular comorbidities (Results Chapter 3).

Ethnicity

WGS data was used to determine the ethnicity of patients recruited to the BRIDGE Study *[BRIDGE core bioinformatics team]*. The majority of patients recruited to the BRIDGE PAH Study were of European ancestry (89 %; Figure 8). The female gender bias was more pronounced amongst non-European patients (3.5 : 1) compared to European patients (2.1 : 1); Fisher's exact test p = 0.022). Non-European patients were more likely to be diagnosed earlier in life with 85 % of non-Europeans classified in the younger group compared to 73 % of Europeans (Fisher's exact test p = 0.004). The ancestry of patients with PAH and PVOD / PCH were similar with 89 % of patients with PAH and 91 % of patients with PVOD / PCH being of European ancestry (Fisher's exact test p = 1).

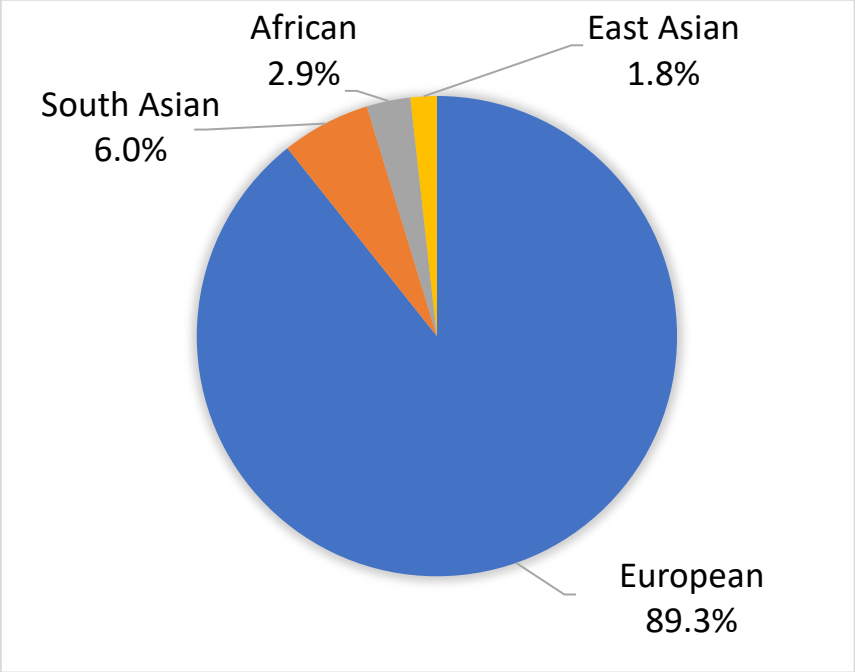


Figure 8. WGS inferred ethnicity of the NIHR BRIDGE PAH Study population

Ethnicity was inferred by comparing 35,114 autosomal SNPs from patients in the study to the same SNPs in 2100 patients of known ethnicity recruited to the 1000 Genomes Project. This was done by Marta Bleda using PC Air. The majority of patients recruited to the study (89 %) were of European ethnicity.

The WGS inferred ethnicity was compared to self-reported ethnicity as collected in the National Health Service and UK National Census (Appendix 13). There were several patients whose self-declared ethnicity was different to their WGS inferred ethnicity. Four of these individuals identified themselves as "British" but were of non-European ethnicity according to the WGS data. Two individuals of European inferred ethnicity identified themselves as "other black" and "Caribbean". These discrepancies may show differences in people's perception of their ethnicity and also the lack of granularity in both the self-reported and WGS inferred categories.

Pulmonary haemodynamics

Pulmonary haemodynamic variables are used to diagnose (mPAP, PCWP and PVR) and risk stratify (RAP, CI and S_vO_2) patients with pulmonary hypertension. Correlation matrices were used to further check the validity of the haemodynamic data. As expected there were significant correlations between haemodynamic variables (Figure 9), with increased right ventricular afterload associated with worse right ventricular function. This has implications

for the selection of covariates in multivariate regression analyses that require independence between the covariates.

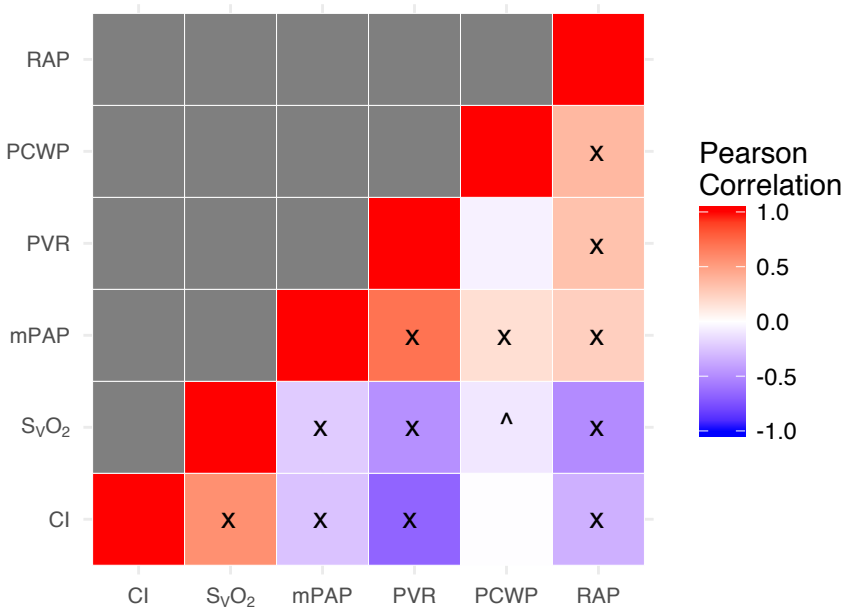


Figure 9. Matrix showing Pearson correlations between haemodynamic variables

Pearson correlation coefficients between the haemodynamic variables collected for all patients recruited to the study (n = 1019) were calculated. A significant correlation is seen between most haemodynamic variables, other than PCWP. Increases in right ventricular afterload (increases in mPAP and PVR) were associated with worse cardiac function (increase in RAP and reductions in CI and S_VO_2).

x – $p < 0.001$, ^ – $p < 0.05$

The distributions and summary statistics (median [IQR]) of the pulmonary haemodynamic variables recorded in the study are summarised in Figure 10. None of these variables had a Gaussian distribution (Shapiro-Wilk test $p \leq 0.001$). The skewed distributions suggest that a small subset of patients have more severe disease with increased right ventricular afterload and worse right ventricular function.

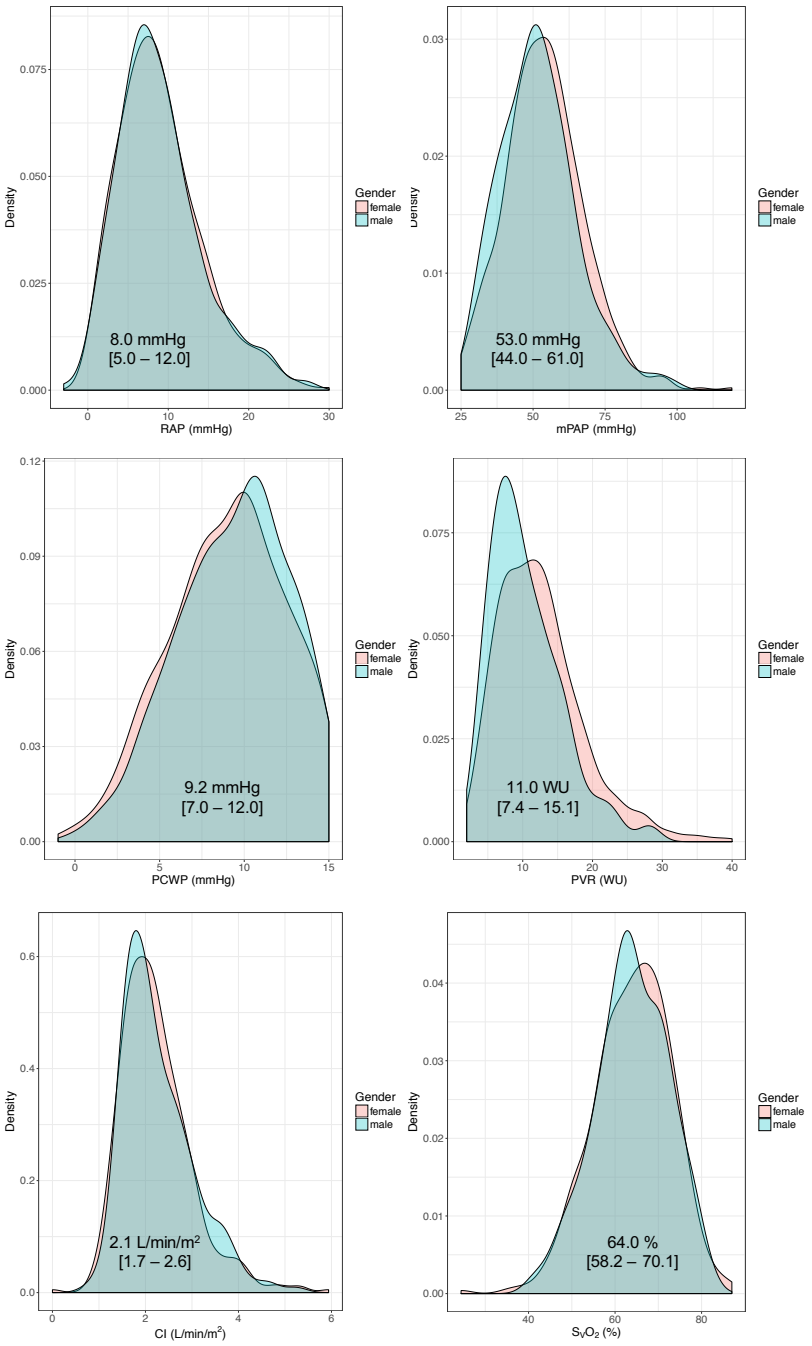


Figure 10. Density plots of the pulmonary haemodynamic variables recorded in OpenClinica

Data was not available for all patients (RAP: n = 876 , mPAP: n = 1003, PCWP: n = 816, PVR: n = 898, CI: n = 436, S_vO_2: n = 657). The skewed distributions of the all the haemodynamic variables are demonstrated. Consequently, rank regression models were used to further analyse the data. The significant difference in PVR between male and female patients demonstrated in the rank regression models can be visualised in the plot.

Median [IQR] of all patients presented.

The distributions of these haemodynamic variables could not be transformed into Gaussian distributions. Therefore, linear regression models could not be used to further assess the data as they make the assumption that the dependent variable is normally distributed. Instead, as discussed in the methods, rank regression models were used to assess associations between haemodynamic variables and both age at diagnosis and gender. Rank regression models use the rank of the observations rather than their actual value.

In a univariate rank regression model, male gender was associated with a significantly lower PVR (regression coefficient [β] = -1.8, standard error [SE] = 0.4, $p < 0.001$). Whereas, older age was also associated with:

lower mPAP (β = -0.2, SE = 0.02, $p < 0.001$),

lower PVR (β = -0.04, SE = 0.01, $p < 0.001$),

lower CI (β = -0.005, SE = 0.002, $p < 0.001$), and

lower S_vO_2 (β = -0.10, SE = 0.02, $p < 0.001$),

as well as a higher PCWP (β = 0.03, SE = 0.01, $p < 0.001$).

In multivariate rank regression models (including age at diagnosis, gender and recruiting centre as covariates), age at diagnosis was significantly associated with all haemodynamic variables except RAP (Table 7). Older age was associated with a lower mPAP and PVR but also reduced cardiac function. In this multivariate model, male gender was still significantly associated with a lower PVR (β = -1.72, SE = 0.39, $p < 0.001$). This can be appreciated from the density plots for PVR that show a clear difference between male and female patients (Figure 10).

The significant association between reduced cardiac function and older age may partly explain the poor prognosis seen in older patients as reported previously and explored later in this chapter.

Table 7. Rank regression models assessing associations between age at diagnosis and haemodynamic variables

	Regression coefficient (β)	**Standard error**	**p**
mPAP	-0.220	0.025	<0.001
PCWP	0.025	0.007	0.001
PVR	-0.041	0.011	<0.001
CI	-0.004	0.002	0.013
S_vO_2	-0.092	0.022	<0.001

Results of multivariate rank regression models using age at diagnosis, gender and recruiting centre as covariates.

CI – cardiac index, mPAP – mean pulmonary artery pressure, PCWP – pulmonary capillary wedge pressure, PVR – pulmonary vascular resistance, S_vO_2 – mixed venous oxygen saturation

Aetiology

In the remainder of this chapter the characteristics of patients with PAH were assessed. The 21 patients with a clinical diagnosis of PVOD / PCH were excluded from this analysis. Initially I used the phenotype data to determine the possible aetiology of PAH in these patients. A history of exposure to drugs and toxins associated with PAH pathogenesis or a family history of the disease were used to categorise patients as drug associated PAH and familial PAH respectively. Where no information was available or there was no history of exposure to drugs or toxins and no family history of PAH, patients were assumed to have idiopathic PAH. I assessed whether there were phenotypic differences between these 3 aetiological groups (familial PAH, drug associated PAH and idiopathic PAH).

Sixty-five patients (7 %) had a family history of PAH (familial PAH), suggesting an underlying genetic cause. Fifty-seven patients (6 %) had a history of exposure to a drug or toxin previously associated with disease pathogenesis (anorexigens, methamphetamines, amphetamines or dasatinib). One patient who had a family history of PAH (an affected sister who was also recruited to the study) and had taken amphetamines was categorised as familial PAH.

The gender bias between these aetiological groups was significantly different (Fisher's exact test $p < 0.001$). Patients with drug associated PAH had the greatest female gender bias (7.0 : 1). Whereas in those with idiopathic PAH the female gender bias was 2.2 : 1 and 1.2 : 1 in familial PAH. The striking gender difference amongst patients with drug associated PAH may be due to an ascertainment bias given that only 110 patients had information of drug and toxin exposure recorded in OpenClinica. Furthermore, the intensity of drug / toxin exposure and its temporal relationship to disease onset was not always accurately recorded. Therefore, misclassification of these patients was a possibility. Although, it can be hypothesised that female patients were more likely to have taken appetite suppressants and therefore were more likely to develop PAH.

Age at diagnosis was significantly different between aetiological groups (Figure 11). Those with familial PAH were diagnosed at an earlier age (37 years [29 - 50]) compared to those with idiopathic PAH (50 years [36 - 64], Dunn's test $p < 0.001$) and those with drug associated PAH

(52 years [41 - 59], Dunn's test $p < 0.001$). This is in keeping with early onset disease in those with a genetic basis for the disease. Yet, even amongst those with familial PAH, 14 patients (21.5 %) were classified in the older group (i.e. over 52 years of age at diagnosis).

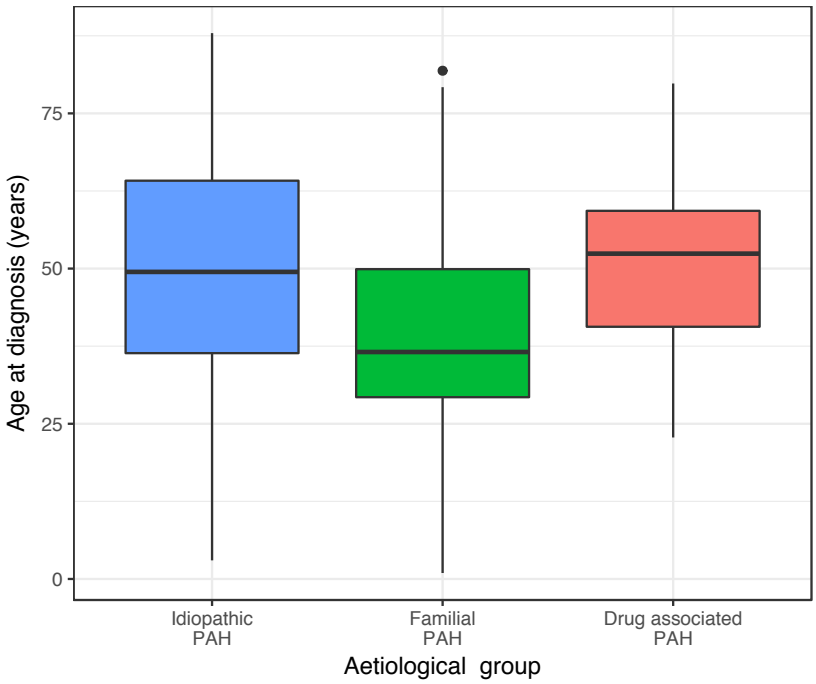


Figure 11. Differences in age at diagnosis between aetiological groups

From the data captured in OpenClinica, patients were classified into aetiological groups based on the presence or absence of a family history of the disease or relevant drug exposure. No genetic data was used to guide this initial classification. The age at diagnosis was not significantly different between the three aetiological groups (idiopathic PAH [n = 877], familial PAH [n = 65] or drug associated PAH [n = 56]).

No significant differences were found between the aetiological groups for the other phenotypic variables following correction for multiple testing. However, this approach is limited by the fact that approximately 17 % of patients with idiopathic PAH are thought to harbour disease-associated variants and are thus misclassified as idiopathic PAH using just clinical data (30).

Genetic basis

To refine this classification further, WGS data was used to identify rare and predicted deleterious variants in genes previously associated with PAH pathogenesis (*BMPR2*, *ACVRL1*, *ENG*, *SMAD1*, *SMAD9*, *KCNK3*, *CAV1*, *TBX4* and biallelic *EIF2AK4*) *[computational genomics*

group including myself]. Amongst patients with a diagnosis of PAH 218 such variants were identified in 204 patients (Table 8). Further assessment of these variants was made, where possible, to identify variants that did not segregate with phenotype within families. The variants identified in the 10 index cases with an affected relative also recruited to the study were also identified in the affected family member. The exception to this was the *SMAD1* variant that did not segregate with phenotype. It was present in the index case but not in their affected son. Consequently, it was considered not to be causal. The 9 patients with biallelic *EIF2AK4* variants (5 variant homozygotes and 4 compound heterozygotes) and a clinical diagnosis of PAH are assessed further in Results Chapter 3 and are excluded from further analysis in this chapter along with the other patients with a clinical diagnosis of PVOD / PCH.

Table 8. Rare and predicted deleterious variants identified in PAH patients

Gene	**n**
BMPR2	159
TBX4	16
ACVRL1	10
Biallelic *EIF2AK4* (number of variants)	9 (14)
ENG	8
SMAD9	6
KCNK3	4
SMAD1*	1
* not segregating with phenotype and therefore not considered associated with PAH	

Ten patients had at least 2 variants in one or more disease associated gene, these patients are summarised in Table 9:

2 patients with 2 *BMPR2* variants,

2 patients with *BMPR2* + *ENG* variants,

2 patients with *BMPR2* + *SMAD9* variants,

1 patient with *BMPR2* + *SMAD1* variants,

1 patient with *TBX4* + *ENG* variants,

2 patients with biallelic *EIF2AK4* + *SMAD9* variants.

Without familial segregation information and functional assessments, it is not possible to be certain whether both variants were causal for the disease, if one variant acted as a disease modifier or even if one or both variants were not associated with disease pathogenesis at all. For classification purposes, genetic and phenotypic considerations were used to identify the most likely "causal" variant for these 10 individuals. The variants identified as most likely to be causal are denoted by an * in Table 9.

1) PTVs were assumed to be the causal variant when it co-occurred with a missense variant.
2) Variants that segregated with phenotype within families were considered causal.
3) There were no features of HHT in the patients with missense *ENG* variants so these *ENG* variants were considered less likely to be causal.
4) Patients with biallelic *EIF2AK4* variants had a characteristic reduced KCO (Results Chapter 3) so these were considered causal.

The identification of the causal variant(s) was more difficult for 3 patients (Table 9). Patient 4 carried variants in both *BMPR2* and *SMAD9* that had previously been associated with the disease (353). The *BMPR2* variant was predicted to result in the loss of a critical cysteine residue in the ligand binding domain of the protein. The *SMAD9* variant resulted in a premature truncation of the protein and is also reported to reduce non-canonical miR signalling similar to known pathogenic *BMPR2* variants (353). However, the *SMAD9* variant is also found in the ExAC database with an allele frequency of $8.24x10^{-6}$. This may suggest that the *SMAD9* variant is not always causal for the disease and additional hits are required to manifest disease. However, the phenotypes of individuals in the ExAC database are not known and therefore further study is required. No relatives were available for a segregation analysis. For categorisation purposes this patient was deemed to have a *BMPR2* variant. Unfortunately, detailed phenotypic information from this patient was not available as they were recruited retrospectively. However, they did have an early age at diagnosis (37 years) and a marked increase in PVR (28 WU) perhaps suggesting more severe disease as a result of multiple hits to the *BMPR2* signalling pathway.

Patients 7 and 8 were confirmed to be half-sisters. Therefore, both *BMPR2* variants were likely to be on the same allele. Furthermore, biallelic *BMPR2* variants are considered fatal in utero.

Table 9. Rare and predicted deleterious variants found in combination

Patient	Gene	Consequence	HGVS	BRIDGE control allele frequency	ExAC allele frequency	CADD score	PolyPhen-2	SIFT
1	*BMPR2**	frameshift variant	c.156_157delTC	Not reported	Not reported	34		
	SMAD1	splice donor variant	c.400+1G>A	Not reported	Not reported	26.2		
2	*BMPR2**	frameshift variant	c.2027_2030dupACCT	Not reported	Not reported	34		
	SMAD9	missense variant	c.1099G>A	Not reported	1.65E-05	27.9	possibly damaging (0.804)	deleterious (0.05)
3	*BMPR2**	deletion (exons 11 – 13)	g.203411230_203418627del	Not reported	NA			
	ENG	missense variant	c.1955G>A	Not reported	Not reported	27.7	probably damaging (1)	deleterious (0)
4	*BMPR2**	missense variant	c.367T>C	Not reported	Not reported	26.6	probably damaging (1)	deleterious (0)
	SMAD9	stop gained	c.880C>T	Not reported	8.24E-06	40		

5	*TBX4**	frameshift variant	c.972delT	Not reported	Not reported	34		
	ENG	missense variant	c.1934G>C	Not reported	3.45E-05	27.9	probably damaging (1)	deleterious (0)
6	*BMPR2**	Deletion (exons 6/7 – 13)	g.203383741_203388057del	Not reported	NA			
	ENG	missense variant & splice region variant	c.1850C>T	Not reported	4.96E-05	32	probably damaging (1)	deleterious (0)
7	*BMPR2*	missense variant	c.1471C>T	Not reported		35	probably damaging (1)	deleterious (0)
	BMPR2	missense variant	c.1772G>A	Not reported		26.5	probably damaging (1)	deleterious (0.02)
8	*BMPR2*	missense variant	c.1471C>T	Not reported		35	probably damaging (1)	deleterious (0)
	BMPR2	missense variant	c.1772G>A	Not reported		26.5	probably damaging (1)	deleterious (0.02)

9	*EIF2AK4**	splice region variant & intron variant	c.257+4A>C	Not reported	8.28E-06	15.5		
	*EIF2AK4**	frameshift variant	c.1392delT	Not reported	2.48E-05	35		
	SMAD9	missense variant	c.419C>T	Not reported	Not reported	24.3	benign (0.216)	deleterious (0.01)
10	*EIF2AK4**	missense & splice region	c.1820T>G	Not reported	8.28E-06	15.5		
	*EIF2AK4**	missense variant	c.2727C>G	Not reported	2.48E-05	35		
	SMAD9	missense variant	c.1099G>A	Not reported	1.65E-05	27.9	possibly damaging (0.804)	deleterious (0.05)

* likely disease "causing" variant (see text). CADD – combined annotation dependent depletion score, ExAC – Exome Aggregation Consortium, HGVS – human genome variation society nomenclature. Biallelic *EIF2AK4* variants described further in Results Chapter 3.

Using WGS-derived information patients were classified as heritable PAH (i.e. carrying a disease-associated variant and/or with a family history of the disease; n = 210), drug associated PAH (n = 49) or idiopathic PAH (n = 730). Eight patients with heritable PAH also had a history of exposure to drugs or toxins associated with PAH. Four of these patients carried variants in *BMPR2* that had previously been associated with PAH, 2 had large deletions in *BMPR2* and 1 had a variant in *TBX4* previously associated with Small Patella Syndrome. One patient with a history of exposure to fenfluramine (time and duration of exposure not known) had a predicted deleterious missense variant in *ENG* (c.1795G>A, p.Ala599Thr). This variant had not previously been associated with HHT (HHT mutation database; http://www.arup.utah.edu/database/ENG/ENG_welcome.php) and the patient did not have any features of HHT either. Therefore, the pathogenicity of this *ENG* variant requires further assessment.

Of interest, in 15 patients with a family history of PAH no rare and predicted deleterious variants (SNV or CNV) were identified in the genes previously associated with disease pathogenesis. This group of patients may prove valuable in the identification of novel genes associated with disease pathogenesis.

I reassessed the phenotypes of patients based on this refined classification (the basis of the current guidelines). The main phenotypic traits of these patients are summarised in Table 10. Patients with heritable PAH were significantly younger at diagnosis, had more severe pulmonary haemodynamic impairment and were less likely to respond to an acute pulmonary artery vasodilator challenge. Furthermore, patients with heritable PAH were more likely to have normal spirometry and a preserved KCO compared to both patients with idiopathic and drug associated PAH. Despite these differences in the severity of disease there were no significant differences in functional class or 6mwt distance between the groups. There were no significant differences between patients with drug associated PAH and idiopathic PAH, other than the increased female preponderance seen in drug associated PAH. More detailed phenotypic assessments of patients with variants in specific genes are provided in Results Chapter 2.

Table 10. Phenotypes of patients based on the current PAH classification

	Heritable PAH	Drug associated PAH	Idiopathic PAH	Corrected p
n [%]	210 [21.2%]	49 [5.0%]	730 [73.8%]	
Ethnicity: Afr / E Asian / Eur / S Asian (n [%])	5 [2.4%] / 4 [1.9%] / 195 [92.9%] / 6 [2.9%]	2 [4.1%] / 2 [4.1%] / 43 [87.8%] / 2 [4.1%]	22 [3.0%] / 12 [1.6%] / 650 [89.0%] / 46 [6.3%]	0.510
Gender: female (n [%])	135 [64.3%]	44 [89.8%]	500 [68.6%]	0.006
Age at diagnosis (years)	40.3 [31.8 – 52.3]	53.6 [40.9 – 59.9]	51.1 [38.1 – 65.6]	<0.001
Drug exposure: Yes (n [%])	8 [3.8%]	49 [100.0%]	0 [0.0%]	<0.001
mPAP (mmHg)	57.0 [49.8 – 66.0]	51.0 [44.2 – 59.0]	51.0 [42.2 – 60.0]	<0.001
PCWP (mmHg)	10.0 [7.0 – 12.0]	8.0 [6.0 – 10.0]	9.0 [7.0 – 11.0]	0.396
CO (L/min)	3.5 [2.8 – 4.3]	4.3 [3.8 – 5.2]	4.0 [3.3 – 5.2]	<0.001
CI (L/min/m^2)	1.9 [1.6 – 2.3]	2.3 [1.9 – 2.7]	2.2 [1.8 – 2.7]	0.002
PVR (WU)	13.7 [9.8 – 18.8]	10.4 [7.4 – 12.3]	10.2 [7.0 – 14.0]	<0.001
S_vO_2 (%)	61.0 [57.0 – 66.6]	64.0 [59.5 – 73.2]	65.5 [59.6 – 71.3]	<0.001
Vasoresponder (n [%])	1 [1.6%]	1 [11.1%]	30 [18.5%]	0.006
FC: 1 / 2 / 3 / 4 (n [%])	3 [1.5%] / 45 [23.2%] / 118 [60.8%] / 28 [14.4%]	0 [0.0%] / 6 [13.0%] / 37 [80.4%] / 3 [6.5%]	15 [2.4%] / 137 [22.0%] / 402 [64.6%] / 68 [10.9%]	0.446
6mwt distance (m)	345.5 [285.8 – 431.0]	260.0 [165.0 – 369.0]	312.0 [159.0 – 411.0]	0.251
KCO (% predicted)	82.0 [72.0 – 94.0]	79.0 [64.0 – 93.2]	70.0 [48.0 – 84.0]	<0.001

Spirometry pattern: Normal / Obstructive / Restrictive (n [%])	90 [74.4%] / 19 [15.7%] / 12 [9.9%]	18 [62.1%] / 9 [31.0%] / 2 [6.9%]	226 [53.2%] / 126 [29.6%] / 73 [17.2%]	0.006

6mwt – six-minute walk test, Afr – African, CI – cardiac index, CO – cardiac output, E Asian – East Asian, Eur – European, FC – functional class, KCO – transfer coefficient for carbon monoxide, mPAP – mean pulmonary artery pressure, PCWP – pulmonary capillary wedge pressure, PVR – pulmonary vascular resistance, S Asian – South Asian, S_vO_2 – mixed venous oxygen saturation. Kruskal-Wallis test used to assess continuous variables, Fisher's exact test used to assess categorical variables, Cochran–Armitage test used to assess ordinal variables.

In a left truncated multivariate Cox proportional-hazards model including age at diagnosis and gender as covariates, disease aetiology was not a significant predictor of survival despite the more severe pulmonary hypertension and worse right ventricular function observed in patients with heritable PAH (Table 11). Age (hazard ratio [95% confidence interval]: 1.037 [1.022 – 1.052]) and gender (1.858 [1.244 – 2.777]) were significant predictors of survival.

Table 11. Prognostic significance of disease aetiology

	Hazard ratio [95% confidence interval]	**p**
Heritable PAH*	0.852 [0.464 – 1.563]	0.604
Drug associated PAH*	1.336 [0.575 – 3.103]	0.501
Age at diagnosis	1.037 [1.022 – 1.052]	< 0.001
Gender (male) ^	1.858 [1.244 – 2.777]	0.002

Results of a multivariate Cox proportional hazards model using disease aetiology, age at diagnosis and gender as covariates.

* compared to patients with idiopathic PAH, ^ compared to female patients

Prognostic variables

Given the apparent lack of a prognostic difference between those with idiopathic and heritable PAH I investigated other variables that may be of prognostic significance. This analysis was restricted to patients with idiopathic PAH (n = 730). Patients with no survival data (date of last contact [for right censoring] or date of death or transplantation; n = 75) were excluded from the analysis. All but four of these excluded patients were recruited from University Hospital Giessen.

The idiopathic PAH subgroup consisted of a mixture of prevalent (n = 449) and incident (n = 206) cases. This included patients who had died prior to the study start date and were recruited retrospectively. In total 155 deaths and 12 transplants were observed. The inclusion of retrospectively recruited and prevalent patients introduces immortal time and survivor biases. Such biases can be minimized in left truncated survival analyses. In left truncated survival analyses 56 events were excluded as they occurred in the unobserved period prior to the study start date. Left truncated survival analyses were performed as secondary analyses due to the small number of events observed.

Initially, phenotypic variables were assessed in univariate Cox proportional hazards models with no left truncation. In this analysis, an event was defined as death, survival defined as time from diagnosis to death and patients were censored if they were transplanted. Variables that were of prognostic significance with a false discovery rate adjusted $p < 0.1$ were taken forward into multivariate Cox proportional hazards models (n = 24; details provided in Appendix 14).

Due to the limitations in data completeness and a relative lack of events, only age at diagnosis, gender and whether patients were incident or prevalent cases were used as covariates in the initial multivariate models. Phenotypic variables that were of prognostic significance following correction for multiple testing are presented in Table 12. In most of these models age at diagnosis, gender and incident / prevalent group were also significant markers of prognosis. The exceptions to this were in the models with either KCO or KCO % predicted, where age at diagnosis was no longer a significant independent prognostic factor.

Table 12. Variables of prognostic significance in multivariate Cox proportional hazards models

	Categorical group	Number of patients	Number of events	Hazard ratio	Confidence interval	Corrected p
Ascites	Present	368	73	5.619	2.497 – 12.646	<0.001
RBBB	Present	256	55	2.470	1.433 – 4.256	0.017
KCO		377	81	0.149	0.065 – 0.339	<0.001
Functional class		621	149	1.514	1.146 – 1.999	0.045
S_aO_2 during RHC		439	93	0.913	0.878 – 0.948	<0.001
Albumin		465	117	0.933	0.898 – 0.968	0.004
S_aO_2 at rest		449	110	0.934	0.904 – 0.965	0.001
S_vO_2		432	95	0.936	0.914 – 0.959	<0.001
RAP		577	135	1.063	1.032 – 1.096	0.001
S_aO_2 pre-walk		437	100	0.939	0.902 – 0.977	0.030
KCO % predicted		366	84	0.967	0.956 – 0.978	<0.001
6mwt distance		549	129	0.996	0.995 – 0.997	<0.001

Results of multivariate Cox proportional hazards models including age at diagnosis, gender and incident / prevalent status as covariates along with the variable being assessed.

6mwt – six-minute walk test, KCO – transfer coefficient for carbon monoxide, RAP – right atrial pressure, RBBB – right bundle branch block, RHC – right heart catheterisation, S_aO_2 – peripheral arterial oxygen saturation, S_vO_2 – mixed venous oxygen saturation.

Phenotypic variables of prognostic significance could be categorised into 4 groups:

1. Those associated with poor RV function (presence of ascites, RBBB, reduced S_vO_2 and increased RAP).
2. Impaired functional assessments (reduced 6mwt distance and higher functional class).
3. Associated with reduced peripheral oxygen saturations (reduced S_aO_2 and reduced KCO).
4. A reduced blood albumin level was also a poor prognostic marker. This may be related to hepatic congestion associated with RV failure but is treated as an additional group.

To identify independent markers of prognosis a larger multivariate Cox proportional hazards model was created using one variable from the four groups described above. Variables were manually selected to maximise the number of patients and events captured in the model, limit confounders and minimize multiple testing by assessing all permutations of variables from the 4 groups. RAP and functional class were selected as these two variables had the most observations recorded in their respective groups. KCO % predicted was selected as S_aO_2 measurements may be confounded by supplemental O_2 use and KCO does not take into account variation due to gender, age and height. In this model 309 patients were included and 73 events were observed. Age at diagnosis, gender, incident/prevalent group, RAP and KCO % predicted were identified as independent prognostic variables (Table 13).

A secondary left truncated survival analysis was performed using a similar protocol to that used in the primary survival analysis to validate the results. A left truncated analysis takes into account the presence of both incident and prevalent patients, minimizing the immortal time bias (where prevalent patients must have already survived a certain period to be recruited into the study). In the left truncated analysis prevalent patients only enter the risk set from the study start date and retrospectively recruited patients (who have died prior to the start of the study) are excluded completely.

Initially, in left truncated univariate survival analyses 11 variables were identified to be of prognostic significance after correcting for multiple testing (Appendix 14). Red cell

distribution width, peripheral oxygen saturations at rest and post 6mwt, as well as KCO % predicted remained of prognostic significance in small multivariate survival models that included age at diagnosis, gender and incident / prevalent group as covariates. The models maintained the proportional hazards assumption and the Cox-Snell residuals were exponentially distributed indicating good fit of the models.

A more complex multivariate Cox proportional hazards model was then created to identify independent prognostic factors using the variables that were most frequently reported: age, gender, incident / prevalent group, S_aO_2 at rest and red cell distribution width. Despite using the most frequently reported variables only 163 patients were included in this left truncated survival analysis and only 34 events were observed. The Cox-Snell residuals in this model were not exponentially distributed suggesting inadequacy of model fit, possibly due to the lack of events observed.

Regardless, the smaller multivariate left truncated survival analyses support the validity of some of the prognostic variables identified in the primary analysis. More data and events are required to assess the significance of the other variables and identify variables of independent prognostic significance.

The significance of KCO as a prognostic factor was assessed further by assessing it in a cohort of patients with normal spirometry (n = 197). The KCO % predicted remained an independent prognostic variable in both the primary analysis (hazard ratio [95% confidence interval]: 0.959 [0.940 - 0.978], $p < 0.001$) and a left truncated analysis (0.970 [0.947 - 0.993], $p = 0.009$).

Table 13. Identification of independent prognostic markers

	Categorical group	Hazard ratio	Confidence interval	Corrected p
Age at diagnosis		1.023	1.003 - 1.043	0.021
Gender	Male	2.936	1.799 - 4.792	<0.001
Incident / Prevalent	Prevalent	0.346	0.180 - 0.665	0.001
RAP		1.059	1.012 - 1.108	0.012
Functional class		1.040	0.666 - 1.625	0.861
KCO % predicted		0.972	0.960 - 0.983	<0.001
Albumin		0.956	0.912 - 1.002	0.060

Results of a multivariate Cox proportional hazards model using the above variables as covariates.

RAP – right atrial pressure, KCO – transfer coefficient for carbon monoxide

Phenotypes of patients categorised by variables of prognostic significance

Patients were categorised based on each of the variables identified to be of prognostic significance in the primary multivariate Cox proportional hazards model. For the continuous variables patients were dichotomised using the median value. All the differences identified between these groups are shown in Appendix 15 and the results are summarised below.

Gender

Forty phenotypic variables were observed to be significantly different between male and female patients after correction for multiple testing. Similar to the entire study cohort described earlier, male patients with idiopathic PAH were significantly older than female patients (median [IQR]: 60 years [42 - 69] vs. 49 years [35 - 62]; Wilcoxon rank-sum test $p < 0.001$). As would be expected in the general population male patients were taller (174 cm [168 - 179] vs. 162 cm [157 - 166]; Wilcoxon rank-sum test $p < 0.001$) and heavier (84 kg [73 - 96] vs. 71 kg [61 - 85]; Wilcoxon rank-sum test $p < 0.001$) than their female counterparts. Male patients had larger lung volumes (FEV_1, FVC, TLC, VA) but a lower KCO compared to women. Although, the % predicted values for these variables showed no difference except for TLC % predicted, which was lower in men (91 % predicted [83 - 99] vs. 97 % predicted [86 - 104];

Wilcoxon rank-sum test p = 0.032). Supporting this observation, a greater proportion of men were reported to have pulmonary fibrosis on CT (11 % vs. 4 %; Fisher's exact test p = 0.049). Furthermore, men had a lower resting S_aO_2 (94 % [91 - 97] vs. 96 % [92 - 98]; Wilcoxon rank-sum test p = 0.008) and desaturated more during the 6mwt (88 % [82 - 94] vs. 93 % [86 - 96]; Wilcoxon rank-sum test p = 0.004).

Male patients had less severe pulmonary haemodynamic impairment compared to females (Wilcoxon rank-sum test p values report):

mPAP:	49 mmHg [41 - 56] vs. 53 mmHg [44 - 62]; p = 0.010,
CO:	4.4 L/min [3.5 - 5.7] vs. 3.9 L/min [3.2 - 4.9]; p < 0.001,
PVR:	8.5 WU [6.3 - 12.5] vs. 11.1 WU [7.4 - 14.8]; p < 0.001.

Despite this there were no significant differences in either their 6mwt distance or functional class.

Significant differences were observed in haematological indices. As may be expected female patients had a lower Hb concentration (148 g/L [134 - 159] vs. 159 g/L [146 - 169]; Wilcoxon rank-sum test p < 0.001), lower haematocrit (0.4 [0.4 - 0.5] vs. 0.5 [0.4 - 0.5]; Wilcoxon rank-sum test p < 0.001) and lower ferritin concentration (64 μg/L [28 - 109] vs. 121 μg/L [60 - 244]; Wilcoxon rank-sum test p < 0.001) compared to male patients. While platelet counts were higher in female patients (242 $x10^9$/l [199 – 294] vs. 197$x10^9$/l [169 – 242]; Wilcoxon rank-sum test p < 0.001).

In addition to the increased prevalence of pulmonary fibrosis, male patients were also more likely to have coronary artery disease (32 [19.4%] vs. 16 [4.4%]; Fisher's exact test p < 0.001) and obesity (15 [9.1%] vs. 10 [2.8%]; Fisher's exact test p < 0.001) compared to females. This may suggest that male patients, who were generally older, may have undiagnosed or mild left ventricular dysfunction contributing to disease pathogenesis. In support of this hypothesis, the left atrial size on echocardiogram (3.9 cm [3.4 - 4.1] vs. 3.5 cm [2.9 - 3.8]; Wilcoxon rank-sum test p = 0.016) and the MRI derived left ventricular end diastolic volume (101 ml [91 - 143] vs. 80 ml [667 - 105]; Wilcoxon rank-sum test p < 0.001) were significantly greater amongst male patients compared to female patients. As this may be influenced by gender differences, indexed values were assessed to investigate this further. Left atrial size index (n

= 106) and left ventricular end diastolic volume index (LVEDVI, n = 111) were calculated by dividing by body surface area where available. LVEDVI was significantly greater in male patients compared to female patients (1.8 ml/m^2 [1.6 - 2.4] vs. 1.7 ml/m^2 [1.3 - 2.1]; Wilcoxon rank-sum test $p_{unadjusted}$ = 0.009). Although there was no significant difference in the left atrial size index based on echocardiographic measurements (0.067 cm/m^2 [0.063 - 0.075] vs. 0.068 cm/m^2 [0.059 - 0.077]; Wilcoxon rank-sum test $p_{unadjusted}$ = 0.835).

Age

When dichotomising idiopathic PAH patients by the median age (51.1 years) significant differences were observed between younger and older patients with respect to pulmonary haemodynamic variables, functional assessments and the presence of comorbidities. Despite younger patients having a higher right ventricular afterload compared to older patients they had better cardiac function and functional assessments (Wilcoxon rank-sum test p value provided unless stated):

mPAP: 54 mmHg [46 - 63] vs. 49 mmHg [41 - 58]; p < 0.001,

PVR: 10.7 WU [7.0 - 15.3] vs. 10.1 WU [7.0 - 13.4]; p < 0.001,

CI: 2.3 L/min/m^2 [1.9 - 2.9] vs. 2.1 L/min/m^2 [1.7 - 2.6]; p = 0.006,

S_vO_2: 68 % [62 - 73] vs. 64 % [58 - 70]; p < 0.001,

NT-ProBNP: 568 ng/L [141 - 1463] vs. 1565 ng/L [360 - 3182]; p = 0.006,

Ankle swelling: 58 [25.2%] vs. 93 [40.4%]; p = 0.004,

Functional class 1/2/3/4: 12 [3.8%] / 95 [30.2%] / 173 [54.9%] / 35 [11.1%] vs. 3 [1.0%] / 42 [13.7%] / 229 [74.8%] / 32 [10.5%]; Cochran–Armitage test p < 0.001,

6mwt distance: 380 m [300 - 438] vs. 218 m [96 - 330]; p < 0.001.

Younger patients were less likely to have cardiovascular diseases, respiratory disorders or abnormal spirometry (Fisher's exact test p provided unless stated):

Diabetes mellitus: 9 [3.5%] vs. 41 [15.5%]; p < 0.001,

Coronary artery disease: 1 [0.4%] vs. 47 [17.8%]; p < 0.001,

Systemic hypertension: 5 [1.9%] vs. 31 [11.7%]; p < 0.001,

Hyperlipidaemia: 2 [0.8%] vs. 28 [10.6%]; p < 0.001,

Valvular heart disease: 0 [0.0%] vs. 7 [2.7%]; p = 0.044,

COPD / interstitial lung disease (ILD):

5 [1.9%] vs. 41 [15.5%]; p < 0.001,

Emphysema on CT (none / minimal / mild / moderate / severe):

157 [96.9%] / 3 [1.9%] / 0 [0.0%] / 2 [1.2%] / 0 [0.0%] vs.

166 [85.6%] / 6 [3.1%] / 12 [6.2%] / 9 [4.6%] / 1 [0.5%];

Cochran–Armitage test p < 0.001,

Fibrosis on CT (none / minimal / mild / moderate / severe):

160 [98.2%] / 2 [1.2%] / 0 [0.0%] / 1 [0.6%] / 0 [0.0%] vs.

176 [90.3%] / 10 [5.1%] / 8 [4.1%] / 0 [0.0%] / 1 [0.5%];

Cochran–Armitage test p = 0.007,

Lung function pattern (normal / obstructive / restrictive):

126 [66.0%] / 24 [12.6%] / 41 [21.5%] vs.

100 [42.7%] / 102 [43.6%] / 32 [13.7%]; p < 0.001.

Given these age differences and the higher proportion of female patients in the younger group (276 [76.0 %] vs. 221 [60.9 %]; Fisher's exact test p < 0.001) I investigated whether phenotypic differences existed between young (age at diagnosis < 51.1 years) male and female patients with idiopathic PAH (n = 363). Young female patients still had a higher PVR (11.6 WU [7.5 - 15.7] vs. 8.4 WU [6.2 - 12.7]; Wilcoxon rank-sum test p = 0.022) compared to male patients but there were no differences in CI or functional class. Despite the lack of difference in cardiac output, the 6mwt distance was lower in female patients (359 m [280 - 426] vs. 437 m [403 - 476]; Wilcoxon rank-sum test p = 0.049).

Figure 12 demonstrates that survival between young male and female patients with idiopathic PAH is similar. Older patients, particularly the older male patients, have a significantly worse survival compared to the younger patients (log rank p < 0.001). However, the presence of cardiovascular risk factors (systemic hypertension, coronary artery disease, diabetes mellitus and hyperlipidaemia), individually or grouped together, was not a significant independent prognostic factor in multivariate Cox proportional hazards models with age at diagnosis and gender included as covariates.

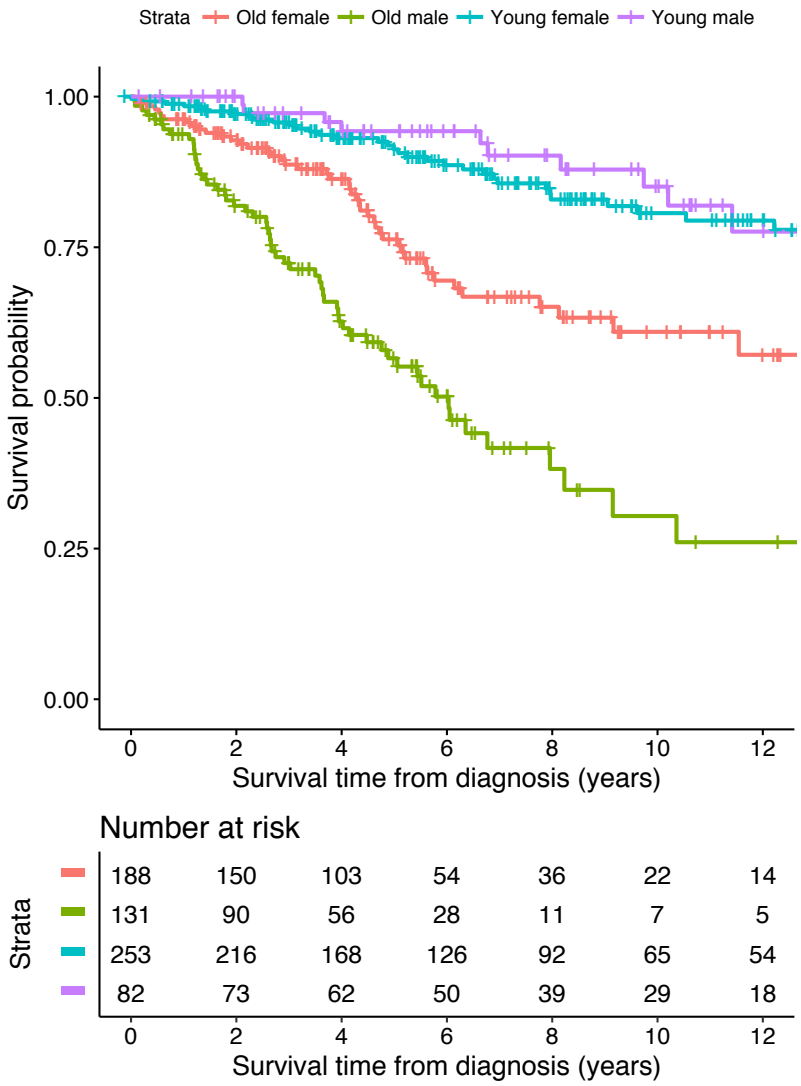


Figure 12. Kaplan-Meier plot demonstrating age and gender differences in survival

The time from diagnosis to death was calculated for patients with idiopathic PAH where survival data was available (n = 655). Patients were censored at the time of last contact or if transplanted. Patients were categorised as young (aged < 51.1 years at diagnosis) or old (age ≥ 51.1 years at diagnosis) and by gender. Younger patients (both male and female) have a better survival compared to older patients (log-rank test $p < 0.001$). Younger male and female patients have a similar prognosis.

Transfer coefficient for carbon monoxide

When dichotomising idiopathic PAH patients by the median KCO % predicted (70 % predicted), patients with a low KCO % predicted were found to be significantly older compared to patients with a high KCO % predicted (65 years [52 - 72] vs. 45 years [35 - 57]; Wilcoxon rank-sum test $p < 0.001$). The proportion of female patients in the low and high KCO % predicted groups

were similar (118 [63.4%] vs. 130 [72.2%] respectively; Fisher exact test p = 0.254). The mPAP was lower in those with a low KCO % predicted (49 mmHg [42 - 56] vs. 55 mmHg [46 - 65]; Wilcoxon rank-sum test p < 0.001) but all other haemodynamic variables were similar between the two groups.

Those with a low KCO % predicted were more likely to have abnormal spirometry, evidence of emphysema on CT and coexisting COPD or interstitial lung disease (ILD).

Lung function pattern (Normal / Obstructive / Restrictive):

90 [48.9%] / 77 [41.8%] / 17 [9.2%] vs.

107 [60.8%] / 32 [18.2%] / 37 [21.0%]; Cochran-Armitage test p < 0.001,

CT emphysema (none / minimal / mild / moderate / severe):

110 [82.7%] / 6 [4.5%] / 10 [7.5%] / 6 [4.5%] / 1 [0.8%] vs.

126 [98.4%] / 1 [0.8%] / 1 [0.8%] / 0 [0.0%] / 0 [0.0%]; Cochran-Armitage test p < 0.001,

COPD / ILD: 34 [18.5%] vs. 5 [2.8%]; Fisher's exact test p < 0.001.

Patients with a low KCO % predicted had lower peripheral oxygen saturations at rest (94 % [90 - 97]) compared to patients with a high KCO % predicted (96 % [94 - 98]; Wilcoxon rank-sum test p < 0.001). Additionally, patients with a low KCO % predicted desaturated more during the 6mwt (87 % [81 - 93]) compared to those with a high KCO % predicted (94 % [89 - 96]; Wilcoxon rank-sum test p < 0.001). The S_vO_2 was also significantly reduced in patients with a low KCO % predicted (64 % [58 - 69]) compared to those with a high KCO % predicted (68 % [61 - 73]; Wilcoxon rank-sum test p = 0.011), although other measures of cardiac function were similar between the two groups.

When restricting the analysis to patients with idiopathic PAH and normal spirometry only 226 patients remained. Patients with a low KCO % predicted had a lower mPAP (48 mmHg [42 - 55] vs. 54 mmHg [46 - 64]; Wilcoxon rank-sum test p = 0.030), lower resting S_aO_2 (95 % [93 - 97] vs. 97 % [95 - 98]; Wilcoxon rank-sum test p = 0.004) and desaturated more during the 6mwt (89 % [81 - 95] vs. 95 % [90 - 97]; Wilcoxon rank-sum test p = 0.008) compared to patients with a high KCO.

Interestingly, a larger proportion of patients in the high KCO % predicted group had a history of congenital heart disease (14 [13.2%]) though these were not thought to be clinically significant, compared to patients with a low KCO % predicted (3 [3.3%]; Fisher's exact test p = 0.020). These included atrial septal defects (n = 10, 1 of which was closed), patent foramen ovale (n = 4) and partial anomalous pulmonary venous drainage (n = 1, this coexisted with an atrial septal defect). In the high KCO % predicted group, patients with congenital heart disease had a higher KCO % predicted (96 % [83 - 101]) compared to those without any evidence of congenital heart disease (84 % [76 - 93]; Wilcoxon rank-sum test p = 0.033). This may suggest there is persistent left to right shunting in patients with congenital heart disease. Therefore, these shunts may be important to disease pathogenesis even though they were deemed clinically insignificant.

Discussion

To the best of my knowledge this is the most comprehensive phenotypic description of patients with idiopathic and heritable PAH. The breadth of information collected as part of the study allowed clinically important observations regarding the heterogeneity of patients with idiopathic PAH to be made. The findings are in keeping with previous reports of the prognostic importance of age at diagnosis, gender, and K_{CO} in this group of patients. Importantly, specific subgroups of patients with mild left ventricular diastolic dysfunction and congenital heart disease with small left to right shunts have been identified. Although these associated conditions may play a role in disease pathogenesis and have implications for treatment and prognosis, these patients were given a clinical diagnosis of idiopathic PAH. Further study of these subgroups is required to definitively ascertain their clinical relevance.

Data quality

As part of the BRIDGE PAH Study, phenotype data from the time of diagnosis was collected from patients diagnosed with idiopathic, heritable and drug- and toxin-associated PAH. The collection of this phenotype data from expert centres in the UK and Europe was the responsibility of research and clinical staff at each site. In order to maximise data capture and quality several steps were put into place at the outset of the study and adapted as the study progressed.

Training and monitoring visits were crucial aspects of the study. Many of the research staff tasked with data entry did not have a clinical background or experience in PH. Therefore, considerable time and effort was put into training research staff at each centre. This is an ongoing process given the turnover of staff. In some centres, this training was inadequate and led to significant errors, such as patients not meeting the inclusion criteria being recruited. Although some errors were identified through the checks put in place it is possible that other errors have gone undetected.

Errors may be systematic in nature or simply relate to mistakes in the transcription of data from clinical records into OpenClinica. Methods to overcome these problems include double data entry by site staff and review of all entries for all patients during monitoring visits (354).

Both of these options would be time consuming and resource intensive given the amount of data captured in study. As a compromise, all data for predefined core variables could be reviewed, in addition to full datasets for randomly selected individuals, during monitoring visits. This would undoubtedly improve the quality of the data entered and reduce missing data.

Missing data could arise due to 1) omission from research staff entering data, 2) incomplete or missing medical records (particularly for older records), or 3) a test not being performed. The lack of complete records for patients reduces the power of the study and introduces selection biases when creating multivariate models that exclude patients with missing data. This is especially true of data that is not missing at random. The amount of missing data could be improved by obtaining further clinical information from the European centres. Imputation of missing data is possible and has been used successfully in similar studies (355).

A significant problem encountered during the monitoring visits was the lack of access to source documentation from legacy patients whose medical records had either been transferred off site, transferred to microfilm or destroyed outright. In the near future, direct transfer of information held in electronic medical records may address some of the issues raised above. Although security and database compatibility issues would need to be addressed.

Whole genome sequencing

A unique aspect of this study was the availability of WGS data for all recruited patients. This allowed the standardised assessment of SNVs and CNVs that may be responsible for disease pathogenesis. Consequently, as shown in the results, it facilitated the accurate categorisation of patients as either idiopathic or heritable PAH. However, the study relied on bioinformatic tools such as CADD, PolyPhen-2 and SIFT, as well as allele frequencies to identify potentially pathogenic variants because of a lack of segregation data and/or experimental evidence of pathogenicity (332-334). A predicted deleterious variant is not necessarily disease causing. Consequently, the reliance on bioinformatic tools may have resulted in the misclassification of patients with predicted deleterious but non-disease-causing variants as heritable PAH. Examples of where this may have occurred are the rare and predicted deleterious missense

variants identified in *ENG*. Most patients with these variants did not have features of HHT, despite the disease being highly penetrant. Furthermore, these variants had not previously been reported in association with HHT or PAH. Despite this concern for over calling pathogenic variants, the proportion of patients that were identified as carrying disease-causing variants was similar to that reported in other studies (42). However, many of these large genetic studies are also limited by the lack of experimental data or familial segregation data.

Prognostic factors

The identification of variables of prognostic significance is subject to survivor bias and immortal time bias. This is a consequence of the inclusion of prevalent patients who were diagnosed prior to the start of the study. In the future, a survival analysis including just incident cases should be possible. However, with approximately 100 new diagnoses of idiopathic PAH in the UK each year it will take a long time to recruit enough patients and observe sufficient numbers of events (deaths) (356).

Left truncated survival analyses, that exclude patients that have died prior to the start of the study and delays entry of prevalent patients into the risk set until after the study start date, can be used to minimize these biases. However, as can be seen in the complex multivariate Cox proportional hazards model, insufficient patient numbers and events were present to properly utilise such statistical techniques.

The identification of the K_{CO} % predicted as a significant prognostic variable in both untruncated and left truncated analyses is of interest. It has previously been identified as a significant prognostic variable in untruncated analyses from single centre cohorts (351, 357). The significance of K_{CO} % predicted remained even in an analysis restricted to patients with normal spirometry. This study supports the phenotypic description of patients with a low D_LCO reported by Trip et al. They also report patients with a low D_LCO were older and more likely to have co-existing cardiovascular diseases (351).

The reason for the poor prognosis associated with a low K_{CO} remains unclear. It is possible that patients with a low K_{CO} have undiagnosed PVOD / PCH or PAH associated with connective

tissue disease. These two forms of PH are also associated with a low K_{CO} and have a worse prognosis compared to idiopathic PAH (300, 358). van der Bruggen et al. showed that patients with a low D_{LCO} had a similar response to pulmonary artery vasodilator therapies compared to patients with a high D_{LCO} (359). The large range observed in K_{CO} may suggest a variety of histopathological changes in the pulmonary circulation of patients with idiopathic PAH. These histopathological changes may be similar to the distinctive changes found in patients with PVOD / PCH and patients with *BMPR2* variants (112, 295).

Summary

As part of BRIDGE PAH Study, I sought to assess the depth and quality of the phenotype data captured in OpenClinica. Along with the bioinformatics team and study co-ordinators, I have set up a system to obtain and verify clinical data, which will form the basis of all subsequent analyses and efforts at new gene discovery. Difficulties were identified in obtaining complete datasets for patients, particularly for those diagnosed at European centres and for those whose clinical records were no longer available. Efforts were made to improve the quality of the data by identifying and checking possible errors with the recruiting centres. Although, without monitoring all data fields for all patients it is impossible to guarantee accurate data.

Survival analyses were performed to identify clinically important phenotypes in patients with idiopathic PAH. Age at diagnosis, gender and KCO were identified as prognostic factors. However, biases in the survival analyses have been identified due to the inclusion of prevalent patients. As more patients are recruited to the study and more survival events are observed these analyses can be repeated to validate these initial findings.

The strengths of this study were the availability of WGS data for all patients and the depth of the phenotype data captured. Patients carrying variants in genes previously associated with PAH were younger at diagnosis and had more severe pulmonary haemodynamic impairment compared to those with idiopathic PAH. These analyses are refined in subsequent results chapters where the phenotypes of patients with variants in specific genes are assessed further.

Results Chapter 2: Phenotype-genotype associations in heritable PAH

Introduction

Heritable PAH is diagnosed in a patient with PAH if they have either a family history of the disease or a pathogenic variant in a gene known to be associated with the disease (1). Ten genes are generally accepted at the present time to be associated with PAH . These are *BMPR2*, *ACVRL1*, *ENG*, *SMAD9*, *SMAD4*, *SMAD1*, *KCNK3*, *CAV1*, *TBX4* and *EIF2AK4*. Other putative genes have been suggested but as yet there is insufficient clinical or experimental evidence to confirm their association with the disease.

Disease associated variants in *BMPR2* account for the majority of cases of heritable PAH. Studies have identified *BMPR2* variants in approximately 17 % of patients with idiopathic PAH and 82 % of patients with heritable PAH (8-16, 30). Patients with *BMPR2* variants are diagnosed at a younger age and have more severe pulmonary haemodynamic impairment. In a large international meta-analysis Evans et al. were able to confirm that *BMPR2* variant carriers had a significantly worse prognosis compared to patients with idiopathic PAH in multivariate models including age and gender as covariates (30). It was suggested that this effect was partly due to more impaired RV function in patients with *BMPR2* variants.

Recently, Trip et al. showed patients with *BMPR2* variants had a higher KCO compared to patients with idiopathic PAH (360). Patients with abnormal spirometry, abnormal lung parenchyma on high resolution computed tomography of the chest, or a significant smoking history were excluded from this study. However, few other studies have sought to characterise *BMPR2* variant carriers despite the fact that *BMPR2* is a widely expressed protein throughout the body and not just expressed in the pulmonary and systemic circulation (361).

Disease associated variants in non-*BMPR2* genes account for approximately 2 % of patients with sporadic PAH (i.e. no family history of the disease) and 7 % of patients with a family history of PAH (235). Variants in *ACVRL1* and *ENG*, which are also associated with HHT, account for the majority of these patients. Most patients with variants in *ACVRL1* or *ENG* have features of HHT. This includes mucocutaneous telangiectasia, recurrent epistaxis and

arteriovenous malformations. The only exception to this being a few cases of childhood onset PAH where these features of HHT may not yet have manifested (11).

Variants in *KCNK3*, *CAV1*, *SMAD1*, *SMAD4*, and *SMAD9* are even more rare and only few patients with these variants have been described (70, 71, 75, 80, 353). As such, it remains unclear if these patients have characteristic phenotypes that differentiate them from patients with idiopathic PAH or patients with variants in other genes. It has been suggested that patients with *KCNK3* variants have worse RV function and do not respond to acute pulmonary artery vasodilator challenges (75, 362). *TBX4* variants were initially associated with Small Patella Syndrome (90). Since then patients with both paediatric and adult onset PAH have also been shown to carry *TBX4* variants (89). All paediatric cases had evidence of Small Patella Syndrome on review. A few patients with adult onset PAH carrying *TBX4* variants have also been described, but they did not show features of Small Patella Syndrome (89, 92). Conversely, screening of a cohort of 23 adult patients with Small Patella Syndrome failed to detect pulmonary hypertension in any of the patients (89). Furthermore, post-mortem histological assessment of one patient with a *TBX4* variant showed changes consistent with PVOD rather than PAH. Thus, the phenotypic consequences of *TBX4* variants appear variable.

As discussed in the previous chapter the strengths of the NIHR BRIDGE PAH Study include its large size, breadth of phenotype data collected and availability of WGS data on all recruited patients. This provides an opportunity to phenotype patients carrying specific disease associated variants. In this chapter I aim to:

1) Assess the burden and consequence of rare and predicted deleterious variants in disease associated genes in PAH patients recruited to the NIHR BRIDGE PAH Study.
2) Describe the characteristics and clinical outcomes of patients with variants in *BMPR2* and other genes previously associated with PAH.

Rare and predicted deleterious variants in genes previously associated with PAH

One-thousand and nineteen patients with PAH had WGS data available and were retained for analysis in the NIHR BRIDGE PAH Study as discussed in Results Chapter 1 and summarised in Figure 5. These patients could be classified as familial PAH (n = 64), drug associated PAH (n = 49), idiopathic PAH (prior to genetic testing; n = 869) or PVOD / PCH (including patients with biallelic *EIF2AK4* variants [Results Chapter 3]; n = 30). The number of patients with rare and predicted deleterious variants in genes associated with PAH are shown in Table 14. Although a higher proportion of patients with a family history of the disease had a variant in a disease associated gene, the greatest number of patients carrying such variants occurred in those with no apparent family history. This may reflect failure to obtain a complete family history, delayed presentation/diagnosis of other family members, reduced penetrance of these variants, or a high frequency of de novo variants.

In Results Chapter 1 no significant differences were identified between patients with idiopathic PAH and drug associated PAH, other than the increased female bias in the latter group. Furthermore, the proportion of patients carrying a variant in a disease associated gene was similar between the two groups. Therefore, in the remainder of this chapter, patients with idiopathic and drug associated PAH (after excluding patients with disease associated variants identified by WGS; n = 779) were treated as a single "PAH" group. These PAH patients were compared to patients with variants in individual PAH associated genes. Patients with PVOD / PCH were again excluded from further analysis.

Table 14. Number of causal variants in genes previously associated with disease pathogenesis identified in each PAH subgroup

Gene	Idiopathic PAH	Familial PAH	Drug associated PAH	PVOD / PCH^
ACVRL1	9 (1%)	1 (2%)	0	0
BMPR2	107 (12%)	45 (70%)	5 (10%)	0
CAV1	0	0	0	0
Biallelic *EIF2AK4*	0	0	0	14 (47%)
ENG	3 (0.3%)	1 (2%)	1 (2%)*	0
KCNK3	4 (0.5%)	0	0	0
SMAD1	0	0	0	0
SMAD4	0	0	0	0
SMAD9	3 (0.3%)	0	0	0
TBX4	13 (1.5%)	2 (3%)	1 (2%)	0

Ten patients also had variants in a second disease associated gene these were discounted for classification purposes as described in Results Chapter 1. *This includes the *ENG* variant of uncertain significance described previously. ^Includes biallelic *EIF2AK4* variant carriers with a clinical diagnosis of idiopathic PAH (Results Chapter 3). Number and percentage of patients in each disease group shown.

BMPR2

Rare and predicted deleterious *BMPR2* variants

In control subjects from other NIHR BRIDGE studies, 24 rare and predicted deleterious variants were identified. Most of these variants (n = 16) occurred in the cytoplasmic tail domain of *BMPR2* (encoded by exons 12 and 13). This included the frameshift variant p.Asn1006Ter, located well after the last PAH associated PTV in *BMPR2* (p.Gln918Ter). All the other variants found in control subjects were missense variants, except for two inframe deletions.

In the NIHR BRIDGE PAH Study 159 rare and predicted deleterious *BMPR2* variants (138 SNVs and 21 CNVs) were identified in 157 patients (two half-sisters both carried 2 rare and

predicted deleterious *BMPR2* missense variants; Results Chapter 1). These variants are listed in Appendix 16. There were 126 unique variants; 37 % of which have previously been reported. The consequence types and locations of these variants are shown in Table 15 and Figure 13. The extracellular domain and serine-threonine kinase domain had a higher proportion of missense variants (p = 0.004), highlighting the importance of key amino acid residues in ligand binding and catalytic activity. Whereas, the smaller proportion of missense variants in the transmembrane domain and cytoplasmic tail suggests that these areas do not carry critical amino acid residues but are of structural importance or required for protein trafficking

Table 15. Consequence types and locations of *BMPR2* variants identified in PAH patients

Consequence type	**ECD**	**TM**	**PK**	**CT**	**NC**	**Total**
Frameshift variant	12 (40%	1 (20%)	15 (28%)	10 (38%)	0 (0%)	**38**
Stop gained variant	5 (17%)	4 (80%)	17 (31%)	12 (46%)	0 (0%)	**38**
Splice region variant	0 (0%)	0 (0%)	0 (0%)	0 (0%)	23 (100%)	**23**
Inframe insertion / deletion	0 (0%)	0 (0%)	2 (4%)	0 (0%)	0 (0%)	**2**
Missense variant	13 (43%)	0 (0%)	20 (37%)	4 (15%)	0 (0%)	**37**
Total	**30**	**5**	**54**	**26**	**23**	**138**

ECD – extracellular domain, TM – transmembrane domain, PK – serine / threonine kinase domain, CT – cytoplasmic tail, NC – non-coding. Percentages calculated with column totals.

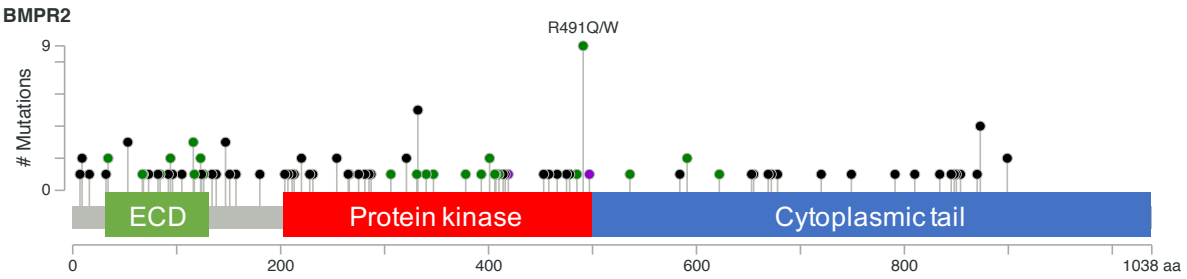


Figure 13. Gene locus plot of *BMPR2* variants identified in PAH patients

Rare (minor allele frequency < 0.0001) and predicted deleterious (CADD score ≥ 15, and PolyPhen-2 or SIFT predictions not classified as "benign" or "tolerated") variants in *BMPR2* were identified in patients recruited to the study. To create the figure the genomic coordinates for each variant, the reference and variant allele at that position, the consequence type of the variant and the amino acid change resulting from the variant were entered into MutationMapper. Splice site variants and deletions are not shown in the figure. The majority of missense variants are located in the extracellular and protein kinase domains where they alter critical amino acid residues resulting in the loss of protein function.

Green – missense variant, Black – protein truncating variant, Purple – inframe variant.

ECD – extracellular domain.

Modified from http://www.cbioportal.org/mutation_mapper.jsp (363)

BMPR2 phenotype - genotype associations

Comparisons were made between patients with *BMPR2* variants (*BMPR2*$^{+/-}$; n = 157) and PAH patients with no variants in disease associated genes (n = 779). All significant phenotypic differences between the two groups are shown in Appendix 17. In summary *BMPR2* variant carriers were significantly younger (39 years [32 - 51] vs. 51 years [38 - 66]; Wilcoxon rank-sum test p < 0.001), and had more severe pulmonary haemodynamic impairment compared to PAH patients (Wilcoxon rank-sum test p values provided):

mPAP: 58 mmHg [51 - 68] vs. 51 mmHg [43 - 60]; p < 0.001,

PVR: 14.2 WU [10.8 - 20.3] vs. 10.2 WU [7.0 - 14.0]; p < 0.001,

CI: 1.8 L/min/m^2 [1.5 - 2.2] vs. 2.2 L/min/m^2 [1.8 - 2.7]; p < 0.001,

S_vO_2: 60 % [56 - 66] vs. 65 % [60 - 71]; p < 0.001.

BMPR2 variant carriers were less likely to have a positive vasoreactive response to an acute pulmonary artery vasodilator challenge (0 [0.0%] vs. 31 [18.1%]; Fisher's exact test p = 0.007). However, functional assessments between the two groups were similar:

6mwt distance: 331 m [287 - 427] vs. 310 m [160 - 410]; Wilcoxon rank-sum test p = 0.210,

Functional class 1/2/3/4: 2 [1.4%] / 32 [21.6%] / 89 [60.1%] / 25 [16.9%] vs. 15 [2.2%] / 143 [21.4%] / 439 [65.7%] / 71 [10.6%]; Cochran-Armitage test p = 0.388.

BMPR2 variant carriers were more likely to have normal spirometry and a preserved KCO compared to PAH patients. They were also less likely to desaturate during a 6mwt.

Lung function pattern (Normal / Obstructive / Restrictive): 81 [81.0%] / 12 [12.0%] / 7 [7.0%] vs. 244 [53.7%] / 135 [29.7%] / 75 [16.5%]; Fisher's exact test p < 0.001,

KCO (% predicted): 82 % [74 - 94] vs. 70 % [49 - 85]; Wilcoxon rank-sum test p < 0.001,

Post walk S_aO_2 (%): 94 % [90 - 97] vs. 91 % [85 - 96]; Wilcoxon rank-sum test p = 0.014.

Significant differences in full blood count indices were observed for the first time between *BMPR2* variant carriers and PAH patients (Wilcoxon rank-sum test p values provided):

Haemoglobin: 162 g/l [152 – 173] vs. 151 g/l [136 – 163]; $p < 0.001$,

Haematocrit: 0.5 [0.5 – 0.5] vs. 0.4 [0.4 – 0.5]; $p < 0.001$,

White blood cell count: 8.8×10^9/l [7.3 – 10.8] vs. 8.1×10^9/l [6.8 – 9.7]; $p < 0.014$.

Other blood test results suggested impaired liver function and possible differences in thyroid function in *BMPR2* variant carriers (Wilcoxon rank-sum test p values provided):

Alanine transferase: 30 iu/l [22 – 42] vs. 23 iu/l [17 – 33]; $p < 0.001$,

Bilirubin: 19 µmol/l [13 – 29] vs. 14 µmol/l [10 – 22]; $p < 0.003$,

Thyroid stimulating hormone: 2.5 mu/l [1.7 – 3.7] vs. 2.0 mu/l [1.2 – 3.1]; $p = 0.013$.

As demonstrated in Results Chapter 1, the age difference between the two groups may in part be responsible for some of the other differences described above. Figure 14 shows a clear difference in the distribution of age at diagnosis between the two groups, with a higher proportion of PAH patients being older than 50 years of age and under 18 years of age.

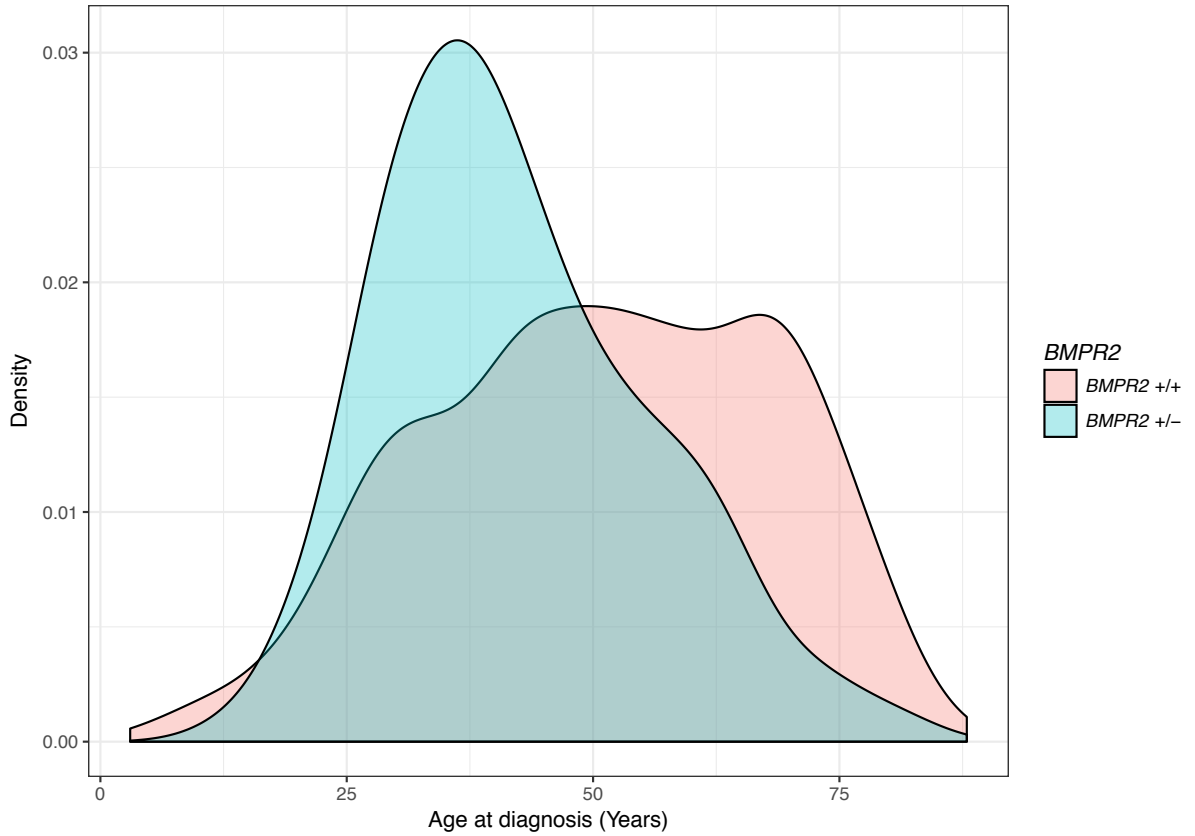


Figure 14. Age distribution of patients with and without *BMPR2* variants

Rare (minor allele frequency < 0.0001) and predicted deleterious (CADD score ≥ 15, and PolyPhen-2 or SIFT predictions not classified as "benign" or "tolerated") variants in *BMPR2* were identified in patients recruited to the study. Comparison was made between patients with PAH and *BMPR2* variants (*BMPR2* +/-; n = 157) and patients with idiopathic PAH (*BMPR2* +/+; i.e. no rare or predicted deleterious variants identified in any of the previously reported PAH associated genes; n = 779). A higher proportion of patients with *BMPR2* variants were in the younger group (age < 51.1 years; 91 %) compared to patients with idiopathic PAH (69 %; Fisher's exact p < 0.001).

To account for this difference the analysis was rerun restricted to patients between the ages of 18 and 50 years. One-hundred and twelve patients with a *BMPR2* variant and 343 PAH patients were included in this analysis. Similar phenotypic differences were identified between the two groups in this restricted analysis. No significant difference was seen in the gender bias between the two groups. Although there was a trend for the female bias to be less pronounced amongst patients with *BMPR2* variants (number of females: 73 [65 %] vs. 267 [78 %] Fisher's exact test p = 0.059). *BMPR2* variant carriers once again demonstrated more severe pulmonary haemodynamic impairment but functional assessments were again similar between the two groups. Similar to the initial analysis, significant differences were

demonstrated in full blood count indices, thyroid stimulating hormone levels and alanine transferase levels. Interestingly, there was no longer a significant difference in KCO between the two groups.

As an alternative analysis, multivariate rank regression models (using *BMPR2* variant status, age at diagnosis, gender and recruiting centre as covariates) were created to assess associations between the presence of rare and deleterious *BMPR2* variants and phenotypic variables (Table 16). These rank regression models support the earlier findings that patients with *BMPR2* variants have more severe pulmonary haemodynamic impairment. It also reinforces the novel associations described between *BMPR2* and blood cell counts. Unlike the previous analysis that excluded patients under 18 years and over 50 years of age, the rank regression model suggests that *BMPR2* does have an independent association with KCO; even when accounting for the age at diagnosis. The lack of a significant association in the previous analysis may be due to a lack of power and/or failure to account for other confounders, such as gender and differences in measurements between recruiting centres. Of note, a significant association was observed between alanine transferase and *BMPR2*. However, assessment of residuals plots revealed that this was being influenced by outliers.

Table 16. Rank regression models assessing the phenotypic differences between patients with idiopathic PAH and those with *BMPR2* variants

Variable	Regression coefficient (β)	Standard error	Corrected p from model
sPAP (mmHg)	8.004	2.108	<0.001
dPAP (mmHg)	5.230	0.986	<0.001
mPAP (mmHg)	5.220	1.164	<0.001
PVR (WU)	4.170	0.499	<0.001
CO (L/min)	-0.841	0.117	<0.001
CI (L/min/m^2)	-0.396	0.069	<0.001
S_vO_2 (%)	-6.132	0.949	<0.001
Hb (g/l)	9.862	2.031	<0.001
WBC (x10^9/l)	0.908	0.252	<0.001
HCT	0.032	0.006	<0.001
KCO	0.148	0.039	<0.001
KCO (% predicted)	10.221	2.644	<0.001

Results of multivariate rank regression models using *BMPR2* status, age at diagnosis, gender and recruiting centre as covariates. Each variable presented in the table was used in turn as the dependent variable. The results of the multivariate rank regression models are presented. Values represent the effect of carrying a *BMPR2* variant.

CI – cardiac index, CO – cardiac output, dPAP – diastolic pulmonary artery pressure, Hb – haemoglobin, WBC – white blood cells, HCT – haematocrit, KCO – transfer coefficient for carbon monoxide, mPAP – mean pulmonary artery pressure, PVR – pulmonary vascular resistance, sPAP – systolic pulmonary artery pressure, S_vO_2 – mixed venous oxygen saturation.

Effect of *BMPR2* variant status on survival

In a univariate Cox proportional hazards model, patients with *BMPR2* variants had a significantly better survival compared to PAH patients (Figure 15; HR [95% confidence interval: 0.561 [0.356 - 0.885]; p = 0.013). This was similar in a left truncated analysis (HR: 0.546 [0.283 - 1.051]; p = 0.070). However, this is not supported by previous studies that

suggest *BMPR2* variant carriers have a worse prognosis (30). In a multivariate model using the variables of prognostic significance identified in Results Chapter 1 as covariates (age at diagnosis, gender, incident/prevalent, KCO % predicted, RAP), the presence of *BMPR2* variants was not a significant prognostic marker (HR: 1.254 [0.660 - 2.384]; p = 0.490).

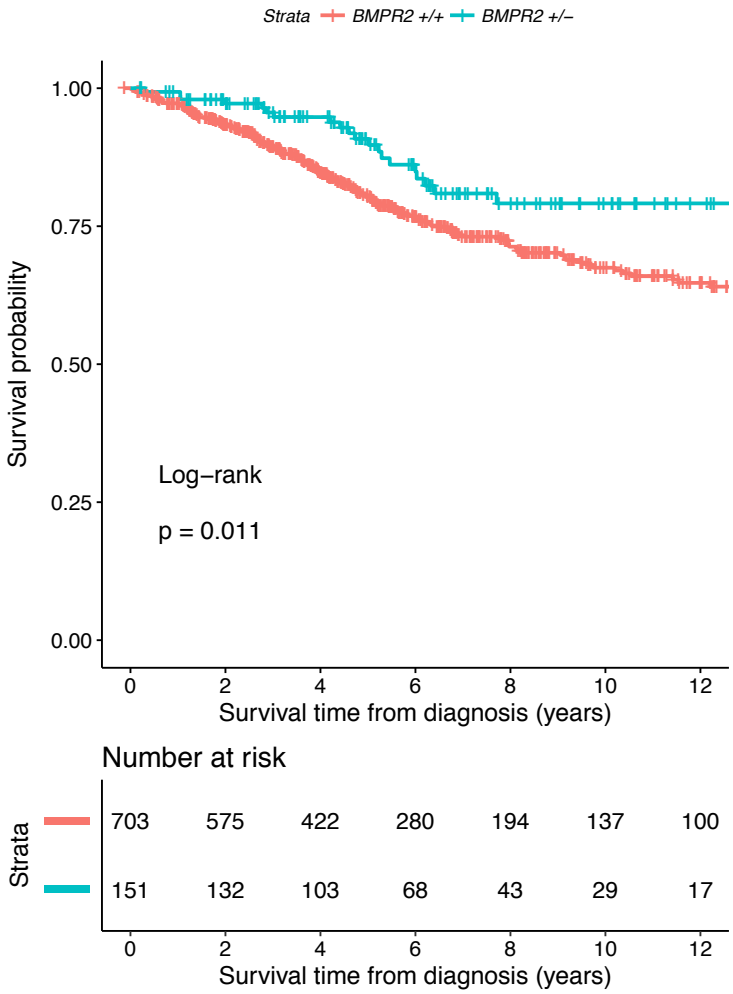


Figure 15. Kaplan-Meier plot demonstrating the effect of *BMPR2* variants on survival

The time from diagnosis to death was calculated where survival data was available. Patients were censored at the time of last contact or if transplanted. Patients were categorised as idiopathic PAH (*BMPR2* +/+ [i.e. not carrying any variants in any of the known disease associated genes]; n = 703) and *BMPR2* variant carriers (*BMPR2*+/- [i.e. carrying a rare and predicted deleterious variant in *BMPR2*]; n = 151). In univariate analysis *BMPR2* variant carriers had a significantly better prognosis compared to patients with idiopathic PAH (log-rank test p = 0.01). However, this does not take into account that patients with *BMPR2* variants were significantly younger and this may have confounded the univariate analysis.

Phenotypic consequences of different *BMPR2* variants

The phenotypic consequences of distinct types of *BMPR2* variant (variant consequence types and location of the variants) were assessed. Initially variants were categorised as deletions (n = 21), PTVs (frameshift and stop gained variants; n = 76), splice site variants (n = 23), and SNVs / indels (missense variants and inframe insertions / deletions < 1 kb; n = 37). After correcting for multiple testing, no differences were observed amongst the different variant consequence types. Although this is in keeping with previous reports, it does not account for confounding by the location of the variant.

To investigate whether there were any *BMPR2* domain specific phenotypes, patients with missense variants and indels were assessed. Patients with PTVs and splice site variants were excluded as the consequence of the variant could not be reliably predicted (for example, did it result in a short protein or undergo nonsense mediated decay). The number of patients retained for this analysis was small, 37. Two variants were identified in the cytoplasmic tail domain, 13 in the extracellular ligand binding domain and 22 in the serine-threonine kinase domain. Again, no significant differences were identified between the different domains. Larger numbers of patients with *BMPR2* variants are required to verify these results.

ACVRL1 and *ENG*

Rare and predicted deleterious *ACVRL1* and *ENG* variants

Sixteen rare and predicted deleterious *ACVRL1* variants and 25 *ENG* variants were identified in control subjects recruited to other NIHR BRIDGE studies. This included an *ACVRL1* stop gained variant (p.Glu159Ter), an *ACVRL1* missense variant (p.Asp254Asn) and 3 *ENG* missense variants (p.Gly214Ser and p.Pro225Leu x2) that had previously been associated with HHT or PAH. Further phenotypic information was unavailable for these 5 subjects.

Amongst patients recruited to the NIHR BRIDGE PAH Study, 10 rare and predicted deleterious *ACVRL1* variants and 8 *ENG* variants were identified (Table 17). Six of these *ACVRL1* variants and just a single *ENG* variant (p.Arg93Ter) had previously been associated with PAH or HHT. Of the 4 remaining *ACVRL1* variants none had been confirmed as being pathogenic in the ARUP Laboratories database (listing HHT associated variants). Moreover, none of the remaining 7 *ENG* variants had previously been listed in the ARUP Laboratory's database either

(364). Of note, other than the stop gained *ENG* variant that had previously been associated with HHT (and was identified in a patient with a history of HHT), all the other *ENG* variants were missense variants. Two of the patients carrying *ENG* missense variants also had large *BMPR2* deletions. While another patient also carried a *TBX4* variant that had previously been associated with Small Patella Syndrome (Table 9).

Phenotypes of patients with rare and predicted deleterious *ACVRL1* and *ENG* variants

Amongst the 794 PAH patients (with no variants in disease associated genes), one patient was reported to have evidence of multiple arteriovenous malformations on CTPA and another had a history of epistaxis, but none had a diagnosis of HHT. Six patients (60 %) with *ACVRL1* variants had a diagnosis of HHT. Only a single patient with an *ENG* variant (p.Arg93Ter) had a diagnosis of HHT. Of the remaining 11 patients with variants in either *ACVRL1* or *ENG* and no diagnosis of HHT, 5 had no mention of arteriovenous malformations, telangiectasia or epistaxis recorded in OpenClinica. The remaining patients were from European centres and did not have further clinical information available.

Phenotypic differences between patients with *ACVRL1* or *ENG* variants and those with PAH were assessed. This study lacks power to identify further phenotypic differences. Additionally, the lack of a clinical diagnosis of HHT in some of the variant carriers and the lack of previously reported associations to PAH or HHT for some variants may suggest some of these variants are not pathogenic. Repeating the analysis with patients carrying a variant validated in the ARUP Laboratories database or with a clinical diagnosis of HHT, restricted the number of *ACVRL1* and *ENG* variant carriers to just 9. No significant differences were observed between these 9 variant carriers and patients with PAH after correcting for multiple testing.

Table 17. Variants in *ACVRL1* and *ENG* identified in PAH patients

Gene	HGVSc	HGVSp	Exon	Consequence type	CADD score	ExAC AF	ARUP database	HHT diagnosis
ACVRL1	c.334C>T	p.Gln112Ter	4 of 10	Stop gained	35		Pathogenic	Yes
ACVRL1	c.647T>G	p.Val216Gly	6 of 10	Missense variant	28.3		Variant of uncertain significance	Yes
ACVRL1	c.1030T>C	p.Cys344Arg	7 of 10	Missense variant	27		Pathogenic	Yes
ACVRL1	c.1450C>G	p.Arg484Gly	10 of 10	Missense variant	27.6		Pending classification	Yes
ACVRL1	c.1451G>T	p.Arg484Leu	10 of 10	Missense variant	32		Pending classification	Yes
ACVRL1	c.1450C>T	p.Arg484Trp	10 of 10	Missense variant	33		Pathogenic	Yes
ACVRL1	c.1451G>A	p.Arg484Gln	10 of 10	Missense variant	33		Pathogenic	No
ACVRL1	c.433C>T	p.Arg145Trp	4 of 10	Missense variant	23.7		Not listed	No
ACVRL1	c.982C>G	p.His328Asp	7 of 10	Missense variant	32		Not listed	No
ACVRL1	c.604G>C	p.Val202Leu	5 of 10	Missense variant	24		Not listed	No
ENG	c.277C>T	p.Arg93Ter	3 of 15	Stop gained	37		Pathogenic	Yes
ENG	c.1934G>C	p.Gly645Ala	15 of 15	Missense variant	27.9	$3.45x10^{-5}$	Not listed	No
ENG	c.1419C>G	p.Ser473Arg	11 of 15	Missense variant	21.6		Not listed	No
ENG	c.1545C>A	p.Asn515Lys	12 of 15	Missense variant	23.4		Not listed	No
ENG	c.10G>A	p.Gly4Ser	1 of 15	Missense variant	22		Not listed	No
ENG	c.1795G>A	p.Ala599Thr	14 of 15	Missense variant	33		Not listed	No
ENG	c.1850C>T	p.Thr617Met	14 of 15	Missense variant	32	$4.96x10^{-5}$	Not listed	No
ENG	c.1955G>A	p.Cys652Tyr	15 of 15	Missense variant	27.7		Not listed	No

TBX4

Rare and predicted deleterious *TBX4* variants

Amongst control subjects recruited to the NIHR BRIDGE Study, 20 variants in *TBX4* were identified. All but 1 of the variants were missense variants. The single frameshift variant (p.Pro412HisfsTer4) occurred in the final exon of the gene and was further downstream than previously reported disease associated PTVs. Furthermore, 9 missense variants also occurred in the final exon. None of these variants have previously been associated with Small Patella Syndrome or PAH (89, 90, 92).

Amongst patients recruited to the NIHR BRIDGE PAH Study 16 *TBX4* variants were identified (2 stop gained, 7 frameshift and 7 missense variants; Appendix 18). All protein truncating variants and all but of the 2 missense variants occurred upstream of the frameshift variant identified in control subjects. One variant had previously been associated with PAH (p.Tyr382Ser) and two with Small Patella Syndrome (p.Pro372SerfsTer14 identified in 2 unrelated patients) (89, 90). Two patients had a family history of PAH. One of these patients with a family history of PAH did not have any other variants in disease associated genes. The second patient had a *TBX4* variant (p.Thr326ProfsTer12) that has previously been associated with Small Patella Syndrome and also a missense variant in the final exon of *ENG* (p.Gly645Ala). This *ENG* variant had not previously been reported in the ARUP Laboratories database and the patient did not have a clinical diagnosis of HHT. This may question the pathogenicity of this *ENG* variant.

Phenotypes of patients with rare and predicted deleterious *TBX4* variants

No significant differences were seen between the *TBX4* variant carriers and patients with PAH. None of the patients were reported to have Small Patella Syndrome. Two patients had paediatric onset disease. None of the patients with *TBX4* variants had a clinical diagnosis of PVOD / PCH or had a reduced K_{CO} recorded (although this was only available for 3 patients).

KCNK3

Rare and predicted deleterious *KCNK3* variants

Fourteen rare and predicted deleterious variants (13 missense variants and 1 inframe insertion) were identified in control subjects recruited to the NIHR BRIDGE Study. None of these variants were previously associated with PAH (78, 92). In the NIHR BRIDGE PAH Study 4 further missense variants were identified in patients with a diagnosis of PAH (Appendix 19). These patients did not carry any other variants in disease associated genes.

Phenotypes of patients with rare and predicted deleterious *KCNK3* variants

None of the patients carrying a *KCNK3* variant had a family history of PAH. Three of the patients were aged under 30 years at the time of diagnosis (carrying the variants p.Ala114Thr, p.Glu182Gln, p.Gly236Ala) and the fourth was 72 years old (p.Arg51Leu). None of the patients had an acute pulmonary artery vasodilator challenge. No significant phenotypic differences were identified between the *KCNK3* variant carriers and patients with PAH.

SMAD9

Rare and predicted deleterious *SMAD9* variants

In NIHR BRIDGE control subjects 21 rare and predicted deleterious *SMAD9* variants (20 missense variants and 1 stop gained variant) were identified. None of these variants have previously been associated with PAH (70, 71, 353). Amongst patients with a diagnosis of PAH 4 missense variants and 1 stop gained variant (p.Arg294Ter – previously associated with PAH) were identified (Appendix 19) (353). Two additional patients carried both biallelic *EIF2AK4* variants and a *SMAD9* missense variant (p.Pro140Leu and p.Gly367Ser) but are not included here as they are classified as PVOD / PCH.

Interestingly, the *SMAD9* stop gained variant was identified in a patient who also carried a previously described missense variant in *BMPR2* (p.Cys123Arg), resulting in the loss of a critical cysteine residue in the extracellular domain (ligand binding domain) of the protein. A second patient also carried a novel missense variant in *SMAD9* (p.Gly367Ser) and a previously described frameshift variant in *BMPR2* (p.Lys678ProfsTer6). For the purposes of the

phenotype-genotype association analyses these 2 patients with variants in both *SMAD9* and *BMPR2* were excluded.

Phenotypes of patients with rare and predicted deleterious *SMAD9* variants

None of the patients carrying a *SMAD9* variant had a family history of PAH. Unsurprisingly, with only 3 patients carrying just a *SMAD9* variant no significant phenotypic differences were seen between the *SMAD9* variant carriers and patients with PAH. These patients did not have any striking phenotypic abnormality. Unlike the 2 patients carrying biallelic *EIF2AK4* variants and a *SMAD9* variant, the other *SMAD9* variant carriers did not have a low KCO. This may suggest that the low KCO was not a phenotypic consequence of the loss of function of *SMAD9*.

As discussed previously in Results Chapter 1, the male patient carrying the *SMAD9* stop gained variant and the *BMPR2* missense variant resulting in the loss of a critical cysteine residue was recruited retrospectively. Therefore, only limited clinical information was available. They had no apparent family history of the disease and were diagnosed at the age of 37 years. At the time of presentation, they had severe pulmonary hypertension (mPAP 76 mmHg, PVR 28.4 WU) and impaired right ventricular function (CI 1.3 L/min/m^2). They were in functional class 3 at diagnosis and only managed 110 m on a 6mwt. The severity of their disease may be associated with two genetic hits to the *BMPR2* signalling pathway.

Discussion

Heritable PAH

A strength of the study was the availability of WGS data for all patients recruited. This allowed an accurate assessment of the burden of rare and predicted deleterious SNVs and CNVs in this cohort of PAH patients. Although next generation sequencing technologies have been used to discover new disease associated genes, such as *CAV1* and *KCNK3*, in families with a history of PAH, no previous studies have used high throughput sequencing in large numbers of patients with both idiopathic and heritable PAH (75, 80). Selection biases are minimized by employing WGS across the entire cohort. It also allowed the identification of several patients carrying variants in more than one disease associated gene. The proportion of patients carrying variants in disease associated genes was similar to what has previously been reported. This supports the validity of the study design and its findings, despite the lack of experimental and familial segregation information.

The *ENG* and *SMAD1* variants occurring in combination are unlikely to pathogenic as suggested by the complementary genetic and phenotypic information available. However, further experimental and clinical work is required to assess the significance of the *SMAD9* variants occurring in combination. In total 4 patients carried rare and predicted deleterious *SMAD9* variants as well as likely causal variants in *BMPR2* or biallelic variants in *EIF2AK4*. In the case of the patient carrying both a protein truncating variant in *SMAD9* and a highly deleterious *BMPR2* ligand binding domain variant, both variants are potentially pathogenic.

BMPR2

One-hundred and fifty-seven patients were found to carry a SNV or CNV in *BMPR2*. This provided sufficient numbers to assess novel phenotype-genotype associations through simple statistical methods and through more complex multivariate modelling that account for potential confounders. The validity of these analyses is supported by the confirmation of a younger age at diagnosis and more severe pulmonary haemodynamic impairment in patients with *BMPR2* variants.

Several novel observations were also identified even when accounting for age, gender and recruiting centre in the multivariate models. Patients with *BMPR2* variants had a significantly higher haemoglobin concentration (162 g/l [152 - 173]) and haematocrit (0.5 [0.5 - 0.5]) compared to patients with PAH (151 g/l [136 - 163] and 0.4 [0.4 - 0.5] respectively). There were no significant differences in markers of iron homeostasis such as the red cell distribution width, ferritin concentration, iron binding capacity or transferrin saturations. However, the study may have lacked power to identify differences in these biomarkers because of missing data. The data completion rates for these biomarkers were 28 %, 28 %, 17 % and 15 % respectively (Appendix 11). Assessment of iron homeostasis is not part of the routine clinical work up of patients with PAH. The mechanism for these observations is unclear. It has been shown, at least *in-vitro*, that *BMPR2* downregulation increases hepcidin, which acts to reduce absorption of iron from the gastrointestinal tract (218). If this were the case, then patients with *BMPR2* variants may be expected to have a lower haemoglobin concentration and haematocrit compared to patients with idiopathic PAH. Conversely, preclinical studies have suggested that *BMPR2* has a redundant role in regulating hepcidin expression but loss of both *BMPR2* and *ActR2a* (another type 2 bone morphogenetic protein receptor) results in a reduced basal expression of hepcidin (365).

White blood cells were noted to be significantly higher in *BMPR2* variant carriers (8.8 $x10^9$/l [7.3 - 10.8]) compared to PAH patients (8.1 $x10^9$/l [6.8 - 9.7]). Previous, clinical studies have not shown any effect of *BMPR2* loss on immune function (112, 115). Although there is *in-vitro* and preclinical evidence for increased inflammatory responses in human pulmonary artery smooth muscle cells and mice with heterozygous *BMPR2* variants (127). Further investigation of the white blood cell differential will be important in assessing the significance of this finding.

One study by Girerd et al. suggested that patients with variants in the cytoplasmic tail domain of *BMPR2* were significantly older and had a reduced right ventricular afterload compared to patients with other *BMPR2* variants (366). It was postulated that activation of the *SMAD* pathway by these cytoplasmic tail domain variants, which was not demonstrated by protein kinase domain variants, might be the cause of this difference. The study also included familial segregation data for many of these variants, supporting their pathogenicity. Another study

reported that the cellular localisation of the BMPR2 receptor varied between variants in different functional domains but no phenotypic comparisons were made (367). In the NIHR BRIDGE PAH Study no phenotypic differences were identified between patients with different variant consequence types or locations. This may be due to a lack of numbers. However, Evans et al., reporting on the largest number of *BMPR2* variant carriers to date, also found no phenotypic differences between variant consequence types.

In a univariate survival analysis *BMPR2* variant carriers had a significantly better prognosis compared to patients with PAH. However, in multivariate analyses including age and gender as covariates, no significant differences were seen. Insufficient observed events (n = 189) at this stage of the study prevented a more conclusive analysis. In comparison, the meta-analysis by Evans et al. that demonstrated a poorer prognosis for *BMPR2* variant carriers, included 448 *BMPR2* variant carriers out of total study population of 1,550, and 354 deaths were observed (30).

Other genes

Despite the size of the study variants in other genes only accounted for 4 % of patients with either idiopathic, drug associated or heritable PAH. Therefore, insufficient numbers were available to identify novel phenotypes amongst patients with variants in these genes. As such the study could only assess previously reported phenotypes. HHT was diagnosed in most patients with *ACVRL1* variants. The lack of such a diagnosis in the majority of patients with *ENG* variants questions the pathogenicity of these variants.

Surprisingly, given the fact it has only relatively recently been described, variants in *TBX4* were the second most commonly identified in patients with PAH. None of the patients in the study were identified as having Small Patella Syndrome, but further specific assessment for this is required, especially in the 2 childhood onset cases.

Summary

The availability of WGS data for all patients recruited to NIHR BRIDGE PAH Study allowed for the unbiased identification of rare and predicted deleterious variants in disease associated genes. As has been reported previously, variants in *BMPR2* were the most common in patients with PAH. Such variants were identified in 12 % of patients with idiopathic PAH and 70 % of those with a family history of PAH. Variants in other genes were much less common. Together they accounted for 4 % of patients with idiopathic PAH and 6 % of patients with a family history of the disease.

Sufficient numbers of patients with *BMPR2* variants were available to perform multivariate rank regression analyses to identify phenotypic differences between these patients and those without any variants in genes previously associated with PAH. As well as validating previous findings of a younger age at disease onset and more severe pulmonary haemodynamic impairment in patients with *BMPR2* variants, several novel phenotype-genotype associations were identified. Patients with *BMPR2* variants had a higher haemoglobin concentration, haematocrit and white blood cell count compared to those with PAH. Further work is required to understand the reasons for these observations and their potential clinical significance.

The small numbers of patients with variants in other genes resulted in a lack of power to identify specific phenotypes associated with these variants. Sixty percent of patients with variants in *ACVRL1* were noted to have HHT. However, only 1 out of 8 patients with *ENG* variants were reported to have HHT. This demonstrated how phenotype data along with additional genetic information can be used to assess the potential pathogenicity of rare and predicted deleterious variants.

Results Chapter 3: Phenotypic consequences of biallelic *EIF2AK4* variants

Introduction

PVOD / PCH is an ultra-rare form of PH that is associated with a poor prognosis. The clinical features described in patients with PVOD / PCH include a low KCO and oxygen desaturation on exertion, as well as the presence of centrilobular ground glass opacification, interlobular septal thickening and mediastinal lymphadenopathy on high resolution computed tomography (HRCT) of the lung parenchyma (300, 308). However, these clinical and radiological features have also been reported in idiopathic PAH (222, 351, 360). Consequently, the clinical distinction between PVOD / PCH and idiopathic PAH can be challenging. It has been estimated that 10 % of patients with PVOD / PCH are misdiagnosed as idiopathic PAH (368, 369). The diagnosis of PVOD / PCH is often only confirmed post mortem or from explanted lungs by histological assessment.

The histological features of PVOD / PCH typically include pulmonary venous obstructions and pulmonary capillary proliferation, although the distribution of these changes within the lung can be heterogeneous (281, 295). Pulmonary artery smooth muscle hypertrophy and intimal hyperplasia, similar to the changes observed in other forms of PAH, may also be present. Furthermore, pulmonary venous changes have been reported in cases of idiopathic PAH, scleroderma-associated PAH and patients with *BMPR2* variants, to varying extents (112, 296).

A major advance in the molecular diagnosis of PVOD / PCH was the finding of biallelic variants in the gene encoding the eukaryotic translation initiation factor 2 alpha kinase 4 (EIF2AK4) in both familial (100 %) and sporadic (20 – 25 %) cases of PVOD / PCH (20, 21). EIF2AK4 is an activator of the integrated stress response pathway, and responds to environmental stresses, including amino acid deprivation, by phosphorylating the alpha subunit of eukaryotic translation initiation factor 2 (281, 288, 370). These discoveries suggest that *EIF2AK4* variants are specific to PVOD / PCH and that finding biallelic *EIF2AK4* variants in a patient with pulmonary hypertension would be diagnostic of PVOD / PCH. Patients with PVOD / PCH have a poor prognosis and risk fatal pulmonary oedema with the use of pulmonary artery

vasodilator therapies (300, 306, 311, 314). Consequently, early and accurate diagnosis is vital to guide clinical management.

Heterozygous variants in the gene encoding the bone morphogenetic protein type 2 receptor (*BMPR2*) are the most common genetic cause of PAH. They are found in approximately 17 % of individuals with idiopathic PAH and 82 % with a family history of the disease (30). Variants in *BMPR2* have also been reported in patients with histologically proven PVOD (41, 284, 285, 300). Thus, there remains considerable uncertainty to what extent the finding of *EIF2AK4* or *BMPR2* variants reliably predicts the clinical phenotype and response to therapy in a population of patients with PAH.

In this chapter I aim to:

1) Assess the burden of *EIF2AK4* variants in the NIHR BRIDGE Study
2) Describe and assess the significance of *EIF2AK4* variants in a heterozygous state
3) Describe the phenotypic, radiological and histological features associated with biallelic *EIF2AK4* variants
4) Assess the clinical outcomes of patients with biallelic *EIF2AK4* variants

Characteristics of patients with a clinical diagnosis of PVOD / PCH

Twenty-one patients (2 %) recruited to the NIHR BRIDGE PAH Study had a clinical diagnosis of PVOD / PCH (20 PVOD and 1 PCH). Just one of these patients had a family history of the disease. One female patient with a clinical diagnosis of PVOD diagnosed at the age of 42 years gave a history of amphetamine use. No other patients gave a history exposure to drugs or toxins associated with PAH. No patients were recorded as having exposure to inorganic solvents. However, exposure to inorganic solvents, whose association with PVOD / PCH was only reported in 2015, was not specifically requested in OpenClinica (294). Furthermore, such exposures are unlikely to have been captured in clinical records prior to this time. The characteristics of these patients are summarised in Table 18.

Patients with a clinical diagnosis of PVOD / PCH had a similar age at diagnosis (54 years [42 - 68]) compared to patients with a diagnosis of PAH (heritable PAH, idiopathic PAH or drug associated PAH; 49 years [35 - 63]; Wilcoxon rank-sum test p = 0.639). The age distribution of patients with a clinical diagnosis of PVOD / PCH shows a bimodal distribution (Figure 7 [Results Chapter 1]) with modes at 46 and 67 years. These modes reflect those with a genetic cause for the disease (biallelic *EIF2AK4* variants) and those without an identifiable genetic cause (idiopathic PVOD / PCH; Figure 16).

A female gender bias (1.9 : 1) was observed in patients with a clinical diagnosis of PVOD / PCH similar to that observed in patients with PAH (2.2 : 1; Fisher's exact test p = 0.892). Haemodynamic measurements were similar between the two groups. However, patients with a clinical diagnosis with PVOD / PCH had a lower KCO and lower O_2 saturations both at rest and on exertion compared to patients with PAH (Wilcoxon rank-sum test $p < 0.05$). Furthermore, their S_vO_2 (57 % [53 - 61]) was lower than in patients with PAH (65 % [58 - 70]; Wilcoxon rank-sum test p = 0.024), despite a similar cardiac output.

The distribution of KCO % predicted is shown in Figure 17. It demonstrates that a substantial proportion of patients (23 %) with a clinical diagnosis of PAH also have a reduced KCO % predicted (< 50 % predicted). This may be due to coexisting parenchymal lung diseases. To exclude this as a potential cause of the reduced KCO % predicted and other differences seen

between patients with PVOD / PCH and PAH, patients with abnormal spirometry were removed from the analysis. Subsequently, just 6 patients with PVOD / PCH and 340 with PAH were retained. No phenotypic characteristics were significantly different between these two groups with normal spirometry. The K_{CO} % predicted of patients with a clinical diagnosis of PVOD / PCH and normal spirometry was 35 % predicted [30 - 40] compared to 76 % predicted [60 - 90] in patients with PAH and normal spirometry. This difference was not of statistical significance after correcting for multiple testing (Wilcoxon rank-sum test $p = 0.070$).

Table 18. Phenotypic differences between patients with a clinical diagnosis of PVOD / PCH and PAH

	PAH	PVOD / PCH	p corrected
n [%]	998 [97.9%]	21 [2.1%]	
WGS ethnicity: African / East-Asian / European / South-Asian (n [%])	29 [2.9%] / 18 [1.8%] / 891 [89.3%] / 60 [6.0%]	1 [4.8%] / 0 [0.0%] / 19 [90.5%] / 1 [4.8%]	1.000
Gender: female (n [%])	683 [68.5%]	13 [61.9%]	0.892
Age at diagnosis (years)	48.8 [35.4 - 62.7]	54.0 [42.1 - 68.1]	0.639
Family history: Yes (n [%])	65 [6.5%]	1 [4.8%]	1.000
Drug exposure: Yes (n [%])	57 [5.7%]	1 [4.8%]	1.000
mPAP (mmHg)	53.0 [44.0 - 61.0]	49.5 [41.8 - 57.2]	0.741
PCWP (mmHg)	9.0 [7.0 - 12.0]	11.0 [8.0 - 12.0]	0.521
CO (L/min)	4.0 [3.2 - 5.0]	3.4 [2.6 - 4.1]	0.323
CI (L/min/m^2)	2.1 [1.7 - 2.6]	1.8 [1.5 - 2.2]	0.320
PVR (WU)	11.0 [7.4 - 15.1]	10.4 [9.2 - 15.3]	0.892
Vasoresponder (n [%])	32 [13.4%]	0 [0.0%]	1.000
Functional class: 1 / 2 / 3 / 4 (n [%])	18 [2.1%] / 189 [21.7%] / 565 [64.9%] / 99 [11.4%]	0 [0.0%] / 2 [10.0%] / 14 [70.0%] / 4 [20.0%]	0.521
KCO (% predicted)	73.0 [54.0 - 88.0]	38.0 [29.6 - 47.0]	<0.001
Resting S_aO_2 (%)	95.0 [92.0 - 97.0]	91.0 [87.0 - 93.5]	0.005
Post walk S_aO_2 (%)	92.0 [85.0 - 96.0]	85.0 [80.0 - 88.0]	0.024
S_vO_2 (%)	64.5 [58.4 - 70.3]	57.1 [52.7 - 61.4]	0.024
6mwt distance (m)	340.0 [220.0 - 420.0]	219.0 [96.0 - 295.0]	0.032

6mwt – six-minute walk test, CI – cardiac index, CO – cardiac output, KCO – transfer coefficient for carbon monoxide, mPAP – mean pulmonary artery pressure, PCWP – pulmonary capillary wedge pressure, PVR – pulmonary vascular resistance, S_aO_2 – peripheral oxygen saturation, S_vO_2 – mixed venous oxygen saturation, WGS – whole genome sequencing. Data presented as median [IQR] unless stated. Wilcoxon rank-sum test used to assess continuous variables, Fisher's exact test used to assess categorical variables, Cochran–Armitage test used to assess ordinal variables.

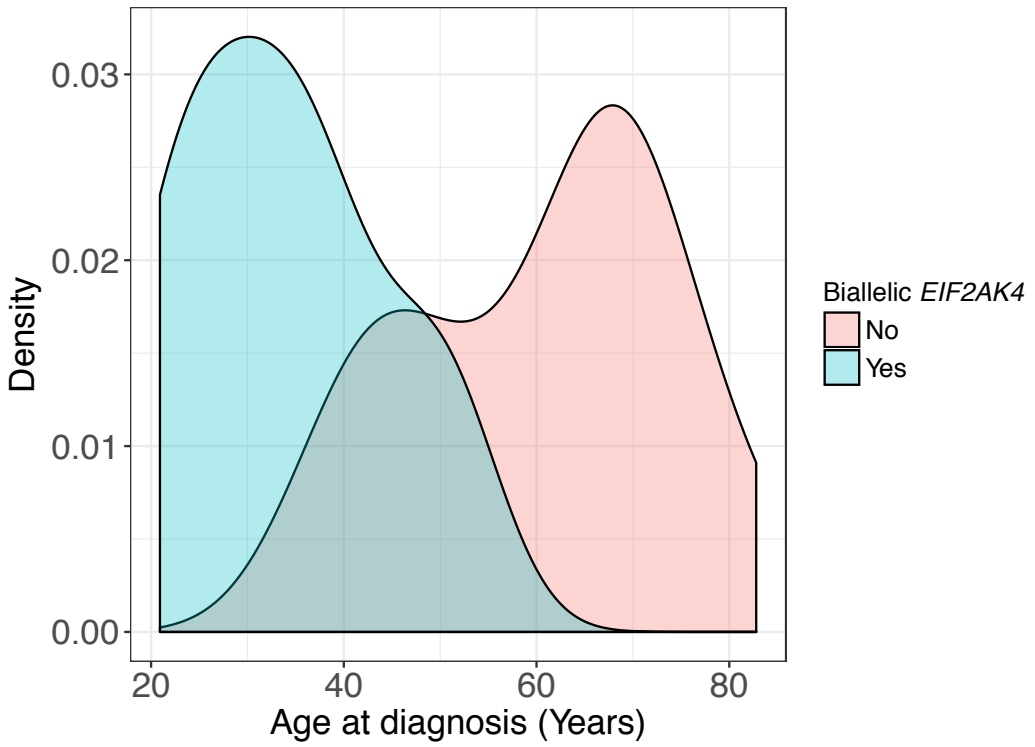


Figure 16. Density plot demonstrating differences in the age at diagnosis in patients with PVOD / PCH

Patients with a clinical diagnosis of PVOD / PCH not carrying biallelic *EIF2AK4* variants (n = 16) were compared to patients carrying rare and predicted deleterious variants in *EIF2AK4* (n = 5). The two groups have visibly different distributions for age at diagnosis. Patients carrying biallelic *EIF2AK4* variants were younger at diagnosis (median [IQR]: 33 years [27 – 39]) compared to those without a mutation (65 years [49 – 69]; Wilcoxon rank-sum test p = 0.002).

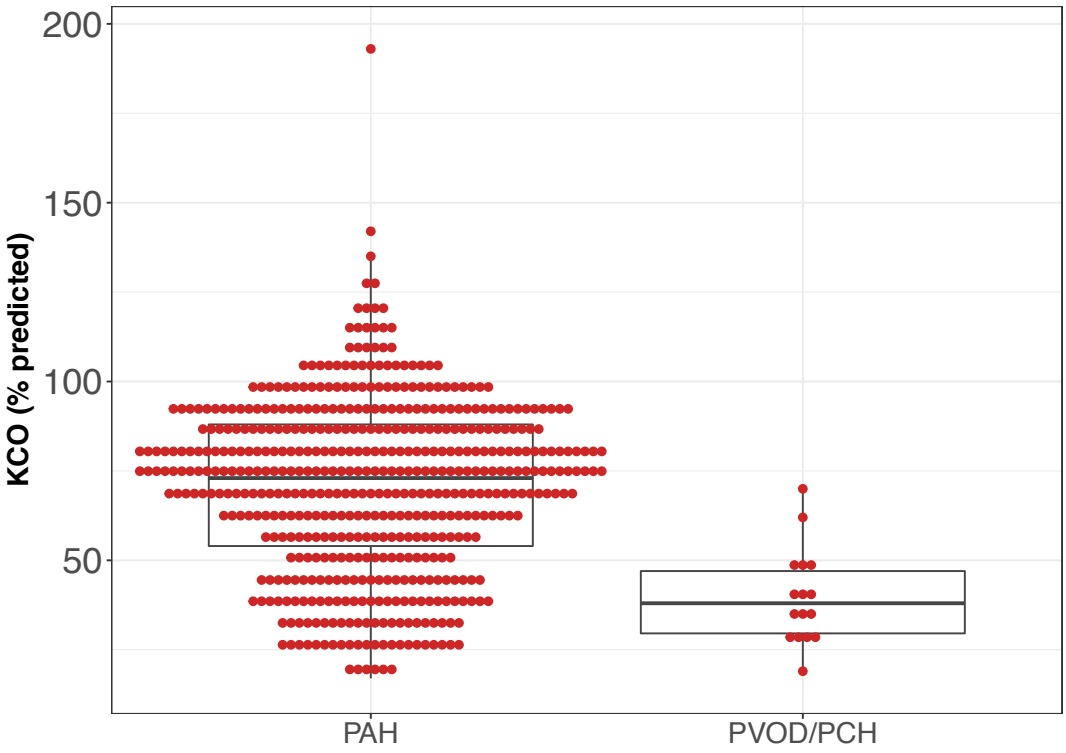


Figure 17. Distribution of KCO % predicted

The KCO % predicted was collected in OpenClinica along with other lung function test results. The figure shows the distribution of KCO % predicted in all patients with a diagnosis of PAH (heritable PAH, idiopathic PAH or drug associated PAH; n = 998) and those with a clinical diagnosis of PVOD / PCH (n = 21). Patients with PVOD / PCH had a low KCO % predicted (all < 75 %). The KCO % predicted was much more variable in patients with PAH but a large proportion (23 %) had a KCO % predicted < 50 %.

EIF2AK4 variants

In the NIHR BRIDGE Study, 74 rare and predicted deleterious *EIF2AK4* variants were identified. Forty-three of these variants (36 in index cases and 7 in relatives) were identified in control subjects recruited to other NIHR BRIDGE studies. Two variants (p.Val385GlyfsTer30 and p.Pro1115Leu) found in control subjects have previously been associated with PVOD / PCH but only when found in a biallelic state (20-22). Importantly, these 43 control subjects were all *EIF2AK4* heterozygotes. The locations of these variants are shown in Figure 18C.

The remaining 31 rare and predicted deleterious *EIF2AK4* variants were identified in patients recruited to the NIHR BRIDGE PAH Study. Fourteen patients carried biallelic *EIF2AK4* variants (Figure 18A and Appendix 20). Seven variants were identified in a homozygous state in 7 patients and 15 variants were identified in a potential compound heterozygous state in 7 patients. Phasing information was not available for these 7 patients with potential compound heterozygous variants. However, as will be discussed later, the phenotypes of these patients were consistent with the variants being in a compound heterozygous state. One patient had two previously reported variants in a compound heterozygous state (p.Arg465ValfsTer38 and c.257+4A>C [splice region variant]) (21, 93). All the other variants identified have not previously been associated with PVOD / PCH.

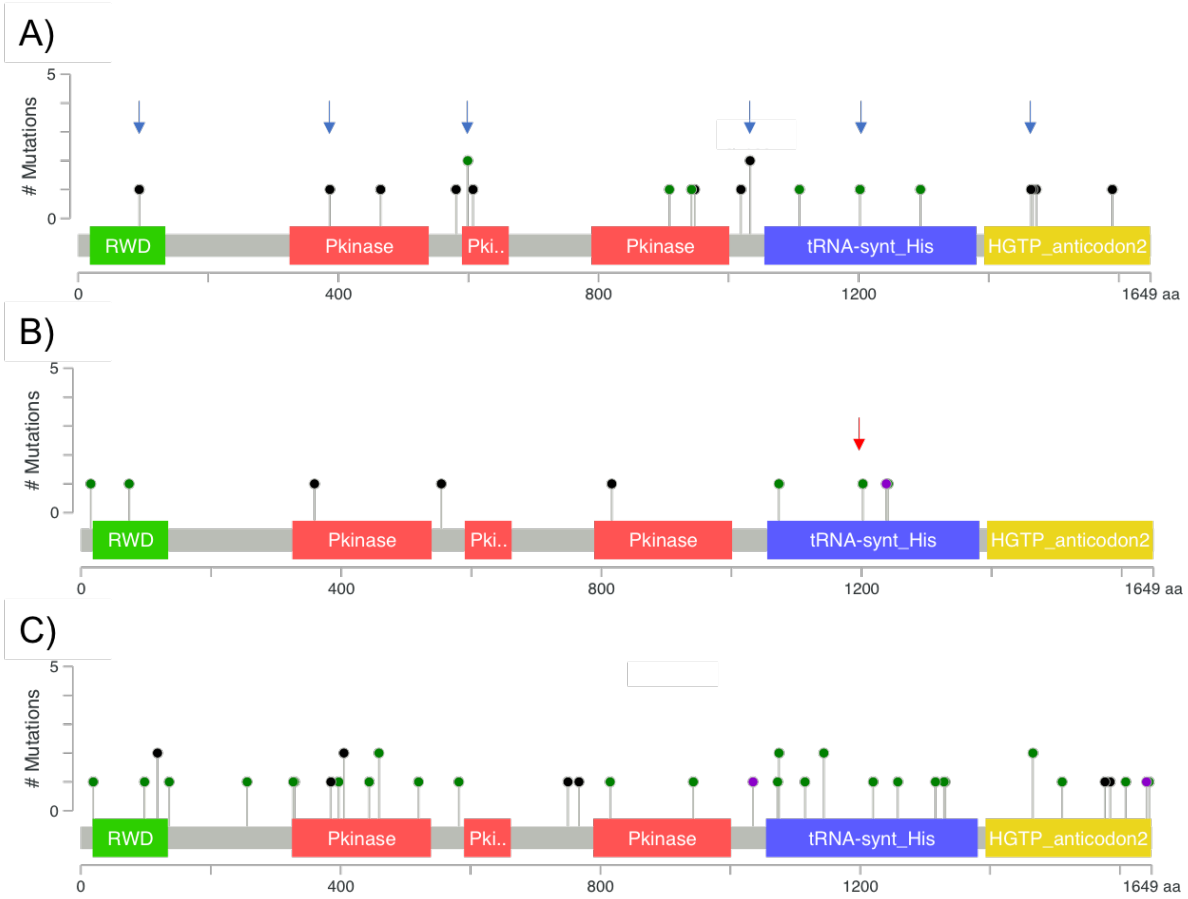


Figure 18. Gene locus plots of rare and predicted deleterious *EIF2AK4* variants identified in the NIHR BRIDGE Study

Rare (minor allele frequency < 0.0001) and predicted deleterious (CADD score ≥ 15, and PolyPhen-2 or SIFT predictions not classified as "benign" or "tolerated") variants were identified in *EIF2AK4* in patients recruited to the NIHR BRIDGE Study. To create the figure the genomic coordinates for each variant, the reference and variant allele at that position, the consequence type of the variant and the amino acid change resulting from the variant were entered into MutationMapper. Splice site variants and deletions are not shown in the figure.

Comparison is made between:

A – variants found in a biallelic state in the NIHR BRIDGE PAH Study,

B – variants found in a heterozygous state in the NIHR BRIDGE PAH Study and

C – variants found in control subjects recruited to other NIHR BRIDGE studies.

Blue arrows – homozygous variants, Red arrow – variant found in a patient with a rare and predicted deleterious *BMPR2* variant.

Significance of *EIF2AK4* variants in a heterozygous state

In addition, 9 *EIF2AK4* variants in a heterozygous state were identified in 8 patients recruited to the NIHR BRIDGE PAH Study (Figure 18B and Appendix 20). One patient had 2 *EIF2AK4* variants (p.Gln816Ter and p.Arg1073Leu), both of which were also identified in her healthy mother, indicating that the 2 variants were carried on the same allele.

The significance of these *EIF2AK4* variants in a heterozygous state is uncertain. Previously, the disease phenotype has only been associated with biallelic *EIF2AK4* variants. Eyries et al. and Best et al. reported that family members of patients with PVOD / PCH and biallelic *EIF2AK4* variants, who themselves carried *EIF2AK4* variants in a heterozygous state, did not show features of PVOD / PCH (20, 21). Similarly, 2 control subjects recruited to other NIHR BRIDGE studies also carried *EIF2AK4* variants in a heterozygous state (p.Val385GlyfsTer30 and p.Pro1115Leu) that have previously been associated with PVOD / PCH when found in a biallelic state.

One patient with an *EIF2AK4* variant in a heterozygous state (p.His1202Tyr) also carried a rare and predicted deleterious variant in *BMPR2* (p.Arg899Ter) that has previously been associated with heritable PAH (Figure 18B and Appendix 20). This patient had no family history of the disease. Family members of this patient had not been recruited to assess whether or not the presence of an *EIF2AK4* variant in a heterozygous state altered the penetrance of the *BMPR2* variant as suggested by Eichstaedt et al. (371).

To further investigate the significance of these *EIF2AK4* variants in a heterozygous state the proportion of patients with such variants and a diagnosis of PAH was compared to the proportion of control subjects in other NIHR BRIDGE studies with *EIF2AK4* variants in a heterozygous state. For this analysis only index cases were included and patients with biallelic *EIF2AK4* variants were excluded. Amongst patients with a diagnosis of PAH 0.8 % carried a rare and predicted deleterious *EIF2AK4* variant in a heterozygous state compared to 0.6 % of control subjects (Fisher's exact test $p = 0.263$).

However, as stated above, disease associated variants when present in a heterozygous state were also found in control subjects. Therefore, the filtering strategy used to identify potentially disease-causing variants may not be as robust in identifying disease-associated (or modifying) variants or those causing a recessive trait. By excluding variants based on their allele frequency in control datasets, variants relevant to disease pathogenesis may be filtered out. To investigate this, an alternative variant filtering strategy was used. In this analysis deleterious variants, regardless of their allele frequency in control datasets, were selected. However, more stringent cut-offs were used to define deleterious variants (CADD score > 20 and *both* SIFT *and* PolyPhen-2 scores predicted to be deleterious or damaging respectively).

Using this modified filtering strategy, 107 predicted deleterious variants in *EIF2AK4* were identified in the NIHR BRIDGE Study. In control subjects recruited to other NIHR BRIDGE studies, 66 *EIF2AK4* variants in a heterozygous state were identified (51 in index cases). Additionally, 2 control subjects carried 2 predicted deleterious variants; potentially in a compound heterozygous state (phasing information was not known; p.Arg166Trp / p.Tyr444Cys and p.Thr943Met / p.Ser1251Cys).

Amongst patients recruited to the NIHR BRIDGE PAH Study, 37 *EIF2AK4* variants were identified. The modified variant filtering strategy identified all 7 *EIF2AK4* homozygotes identified using the original variant filtering strategy. However, the modified variant filtering strategy also identified a new patient with a clinical diagnosis of PVOD who had 2 predicted deleterious PTVs in a potential compound heterozygous state (p.Tyr907Ter and p.Glu184GlyfsTer13). These 2 variants were not retained by the original variant filtering strategy as they had allele frequencies of 0.000106 and 0.00014 in the UK10K study. Conversely, the modified variant filtering strategy did not retain a splice site variant c.257+4A>C (as it had a CADD score of just 15.5) that was found in a patient with a clinical diagnosis of PAH, a low K_{CO} % predicted (33 %) and who also had a second PTV in *EIF2AK4*. Furthermore, this splice site variant when found in a homozygous state has previously been associated with PAH (93).

With the modified variant filtering strategy 15 *EIF2AK4* variants were identified in a heterozygous state in 14 patients with a clinical diagnosis of PAH. Two of these variants had

previously been associated with PVOD / PCH (p.Arg465ValfsTer38 and p.Arg1256Ter) (20, 21). In an overrepresentation analysis limited to index cases, 1.5 % of patients with a clinical diagnosis of PAH carried an *EIF2AK4* variant in a heterozygous state compared to 0.8 % of controls (Fisher's exact test p = 0.059). Therefore, there remains insufficient evidence based on the above analyses alone to suggest a role for *EIF2AK4* variants in a heterozygous state in PAH pathogenesis.

The differences between the two strategies highlights the difficulties relying on computational predictions alone in identifying disease associated variants. Neither variant filtering strategy could be considered superior. Additional phenotypic, experimental or familial segregation data, could provide additional evidence when identifying disease associated / modifying variants.

Both patients with *EIF2AK4* variants identified in a potential compound heterozygous state with just one variant filtering strategy, have features to suggest these variants are disease associated (a clinical diagnosis of PVOD and / or a low KCO). Therefore, in the phenotype-genotype association analyses in the next section both patients were considered to have biallelic *EIF2AK4* variants. However, only *EIF2AK4* variants in a heterozygous state identified using the original filtering strategy were taken forward.

EIF2AK4 phenotype-genotype associations

PVOD

Amongst patients with a clinical diagnosis of PVOD / PCH (n = 21), 6 patients (29 %) carried biallelic *EIF2AK4* variants. None of the patients with a clinical diagnosis of PVOD / PCH carried rare and predicted deleterious *BMPR2* variants or *EIF2AK4* variants in a heterozygous state.

No significant differences were observed between patients with idiopathic PVOD / PCH and patients with PVOD / PCH carrying biallelic *EIF2AK4* variants after correcting for multiple testing. A trend towards a younger age at diagnosis amongst those with biallelic *EIF2AK4* variants was observed (36 years [28 - 42] vs. 67 years [52 - 69]; Wilcoxon rank-sum test p = 0.073). KCO % predicted was recorded in 16 patients with a diagnosis of PVOD / PCH, all of

whom had a reduced KCO % predicted but there was no difference between the two groups (biallelic *EIF2AK4*: 32 % predicted [23 – 36] vs. idiopathic: 40 % predicted [34 – 50]; Wilcoxon rank-sum p = 1.000). Haemodynamic parameters and functional assessments were similar between the two groups (biallelic *EIF2AK4* vs. idiopathic; Wilcoxon rank-sum p value provided unless stated):

mPAP (mmHg):	46 [38 – 54] vs. 50 [46 – 58];	p = 1.000,
PVR (WU):	9.8 [9.3 – 10.2] vs. 12.6 [9.2 – 15.4];	p = 1.000,
CI (L/min/m^2):	2.2 [2.0 – 2.3] vs. 1.8 [1.4 – 2.0];	p = 1.000,
6mwt distance (m):	183 [96 – 274] vs. 219 [96 – 295];	p = 1.000,
Functional class (1/2/3/4):	0 [0%] / 2 [33%] / 1 [17%] / 3 [50%] vs.	
	0 [0%] / 0 [0%] / 13 [93%] / 1 [7%];	Cochran-Armitage
	p = 1.000.	

PAH

Unexpectedly, 9 patients with a clinical diagnosis of PAH (8 with idiopathic PAH and 1 with a family history of PAH) carried biallelic *EIF2AK4* variants. In these patients there was no clinical suspicion of PVOD / PCH.

For the subsequent phenotype-genotype association assessment patients with a clinical diagnosis of PAH were divided into 4 groups based on their genotype (*BMPR2* variant carriers [n = 157], biallelic *EIF2AK4* variant carriers [n = 9], heterozygous *EIF2AK4* variant carriers [n = 7] and those with idiopathic PAH [no family history of PAH and no variants in genes previously associated with the disease; n = 772]). The patient with both a *BMPR2* variant and an *EIF2AK4* variant in a heterozygous state was classified in the *BMPR2* group. Fifty-three patients with variants in other PAH associated genes or a family history of the disease were excluded as were the patients with a clinical diagnosis of PVOD / PCH irrespective of their genotype. Phenotypic comparisons were made between the 4 groups (Table 19). Significant differences attributable to patients with *EIF2AK4* variants in a biallelic or heterozygous state are summarised below.

Patients with biallelic *EIF2AK4* variants and a clinical diagnosis of PAH were significantly younger (29 years [23 - 38]) compared to patients with idiopathic PAH (52 years [38 - 66]; Dunn's test p = 0.010). No differences in haemodynamic or functional assessments were seen between patients with biallelic *EIF2AK4* variants and patients with *BMPR2* variants or patients with idiopathic PAH. There was a trend towards a higher S_vO_2 in patients with biallelic *EIF2AK4* variants (79 % [60 - 73]) compared to patients with *BMPR2* variants (60 % [56 - 66]; Dunn's test p = 0.075). Despite the preserved cardiac function in patients with biallelic *EIF2AK4* variants, a higher proportion (56 %) presented with syncope at diagnosis compared to patients with *BMPR2* variants (40 %) or idiopathic PAH (27 % ; Fisher's exact test p = 0.046).

Patients with clinical diagnosis of PAH and biallelic *EIF2AK4* variants had a significantly lower KCO (33 % predicted [30 - 35]) compared to patients with *BMPR2* variants (82 % predicted [74 - 94]; Dunn's test p < 0.001) and idiopathic PAH (70 % predicted [49 - 85] ; Dunn's test p = 0.001). In support of this finding patients with biallelic *EIF2AK4* variants also desaturated more during a 6mwt (76 % [74 - 82]) compared to patients with *BMPR2* variants (94 % [90 - 97]; Dunn's test p < 0.001) and idiopathic PAH (91 % [85 - 96]; Dunn's test p = 0.003). Furthermore, the haemoglobin concentration in patients with biallelic *EIF2AK4* variants (178 g/l [165- 183]) was higher than those with idiopathic PAH (151 g/l [136 - 163]; Dunn's test p < 0.001).

Although, no significant differences were seen in FEV_1, FVC or TLC a higher proportion of patients with biallelic *EIF2AK4* variants had a normal spirometry pattern (75%) compared to those with idiopathic PAH (54%, Fisher's exact test p < 0.001). Patients with biallelic *EIF2AK4* variants were also more likely to have digital clubbing (43 %) compared to patients with *BMPR2* variants (11 %) or idiopathic PAH (4 %; Fisher's exact test p = 0.005).

Similar results were observed after limiting the analysis to patients with normal spirometry. Patients with biallelic *EIF2AK4* variants had a significantly lower KCO (32 % predicted [29 - 38]) compared to both *BMPR2* variant carriers (82 % predicted [72 - 94]; Dunn's test p < 0.001) and those with idiopathic PAH (72 % predicted [55 - 89]; Dunn's test p = 0.001).

No significant differences were observed between patients with *EIF2AK4* variants in a heterozygous state and idiopathic PAH patients.

Amongst patients with a clinical diagnosis of PAH who had a KCO recorded, 114 (23 %) had a KCO less than 50 % predicted. Most of these patients (84 %) were over the age of 50 years at diagnosis. Only 13 patients aged under 50 years at diagnosis and who had normal spirometry had a KCO less than 50 % predicted. Six of these 13 patients carried biallelic rare and predicted deleterious *EIF2AK4* variants. Therefore, the sensitivity and specificity of a low KCO (< 50 % predicted) and early age at diagnosis (< 50 years) for identifying biallelic *EIF2AK4* variant carriers was 0.86 and 0.98 respectively. However, the positive predictive value was just 0.50. Nevertheless, in terms of the diagnostic yield, while genetic testing for biallelic *EIF2AK4* variants in the entire cohort of patients diagnosed clinically with PAH yielded a 1 % detection rate, the presence of biallelic *EIF2AK4* variants in PAH patients with a KCO < 50 %, normal spirometry and aged under 50 at diagnosis was 50 %.

Patients with a clinical diagnosis of PAH and biallelic *EIF2AK4* variants had a similar phenotype (low KCO and early age at diagnosis) when compared to patients with a clinical diagnosis of PVOD / PCH and biallelic *EIF2AK4* variants (Figure 19). However, while 43 % of patients with a clinical diagnosis of PAH and biallelic *EIF2AK4* variants had digital clubbing none of the patients with a clinical diagnosis of PVOD / PCH and biallelic *EIF2AK4* variants were clubbed.

Table 19. Phenotypic differences by genotype in patients with a clinical diagnosis of PAH

	Idiopathic PAH	***BMPR2***	**Heterozygous *EIF2AK4***	**Biallelic *EIF2AK4***	**p corrected**
n [%]	772 [81.7%]	157 [16.6%]	7 [0.7%]	9 [1.0%]	
WGS population: African / East-Asian / European / South-Asian (n [%])	24 [3.1%] / 13 [1.7%] / 687 [89.0%] / 48 [6.2%]	3 [1.9%] / 2 [1.3%] / 147 [93.6%] / 5 [3.2%]	0 [0.0%] / 1 [14.3%] / 6 [85.7%] / 0 [0.0%]	0 [0.0%] / 0 [0.0%] / 3 [33.3%] / 6 [66.7%]	<0.001
Gender: female (n [%])	538 [69.8%]	105 [66.9%]	6 [85.7%]	4 [44.4%]	0.507
Age at diagnosis (years)	51.5 [38.2 - 65.5]	39.3 [32.0 - 51.1]	39.8 [32.0 - 64.3]	29.3 [23.0 - 37.6]	<0.001
Family history: Yes (n [%])	0 [0.0%]	45 [28.7%]	0 [0.0%]	1 [11.1%]	<0.001
mPAP (mmHg)	51.0 [43.0 - 60.0]	58.0 [51.0 - 68.2]	43.0 [42.0 - 54.0]	52.0 [46.0 - 65.0]	<0.001
PCWP (mmHg)	9.0 [7.0 - 11.0]	10.0 [7.0 - 12.0]	10.0 [8.0 - 11.0]	11.0 [7.5 - 12.5]	0.802
CI (L/min/m^2)	2.2 [1.8 - 2.7]	1.8 [1.5 - 2.2]	2.2 [1.8 - 2.8]	2.9 [1.7 - 3.1]	<0.001
S_vO_2 (%)	65.4 [59.5 - 71.3]	60.1 [56.0 - 65.9]	63.5 [60.0 - 72.0]	69.5 [60.4 - 73.2]	<0.001
PVR (WU)	10.3 [7.0 - 14.0]	14.2 [10.8 - 20.3]	7.7 [6.0 - 9.3]	8.6 [7.5 - 13.1]	<0.001
Vasoresponder: (n [%])	31 [18.3%]	0 [0.0%]	0 [0.0%]	0 [0.0%]	0.01
Functional class: 1 / 2 / 3 / 4 (n [%])	15 [2.3%] / 140 [21.2%] / 435 [65.8%] / 71 [10.7%]	2 [1.4%] / 32 [21.6%] / 89 [60.1%] / 25 [16.9%]	0 [0.0%] / 3 [42.9%] / 4 [57.1%] / 0 [0.0%]	0 [0.0%] / 1 [11.1%] / 8 [88.9%] / 0 [0.0%]	0.447

6mwt distance (m)	309.5 [157.0 - 409.5]	331.0 [286.5 - 426.5]	370.0 [323.0 - 453.0]	502.0 [350.0 - 527.0]	0.226
Post walk S_aO_2 (%)	91.0 [85.0 - 96.0]	94.0 [89.8 - 97.0]	94.0 [80.0 - 96.0]	76.0 [74.0 - 82.0]	0.001
HCT	0.4 [0.4 - 0.5]	0.5 [0.5 - 0.5]	0.5 [0.5 - 0.5]	0.5 [0.5 - 0.5]	<0.001
Hb (g/l)	150.5 [136.0 - 163.0]	162.0 [152.0 - 173.2]	170.5 [157.7 - 171.8]	178.0 [165.0 - 182.8]	<0.001
FEV_1 (L)	2.2 [1.8 - 2.8]	2.6 [2.2 - 3.3]	2.4 [1.9 - 3.0]	3.5 [2.3 - 4.1]	<0.001
FVC (L)	3.0 [2.4 - 3.8]	3.5 [2.9 - 4.2]	3.2 [2.2 - 3.5]	4.1 [3.1 - 5.2]	<0.001
KCO (% predicted)	70.0 [49.0 - 85.3]	82.0 [74.0 - 94.0]	81.0 [72.0 - 95.0]	33.0 [30.2 - 34.7]	<0.001
Spirometry pattern: Normal / Obstructive / Restrictive (n [%])	241 [53.7%] / 135 [30.1%] / 73 [16.3%]	81 [81.0%] / 12 [12.0%] / 7 [7.0%]	3 [60.0%] / 0 [0.0%] / 2 [40.0%]	6 [75.0%] / 2 [25.0%] / 0 [0.0%]	<0.001
Clubbing: Yes (n [%])	11 [3.7%]	7 [10.6%]	0 [0.0%]	3 [42.9%]	0.005
Syncope at onset: Yes (n [%])	131 [27.0%]	41 [39.8%]	0 [0.0%]	5 [55.6%]	0.046

6mwt – six-minute walk test, CI – cardiac index, FEV_1 – forced expiratory volume in 1 second, FVC – forced vital capacity, Hb – haemoglobin, HCT – haematocrit, KCO – transfer coefficient for carbon monoxide, mPAP – mean pulmonary artery pressure, PCWP – pulmonary capillary wedge pressure, PVR – pulmonary vascular resistance, S_aO_2 – peripheral arterial oxygen saturation, S_vO_2 – mixed venous oxygen saturation, WGS – whole genome sequencing. Data presented as median [IQR] unless stated. Kruskal-Wallis test used to assess continuous variables, Fisher's exact test used to assess categorical variables, Cochran–Armitage test used to assess ordinal variables.

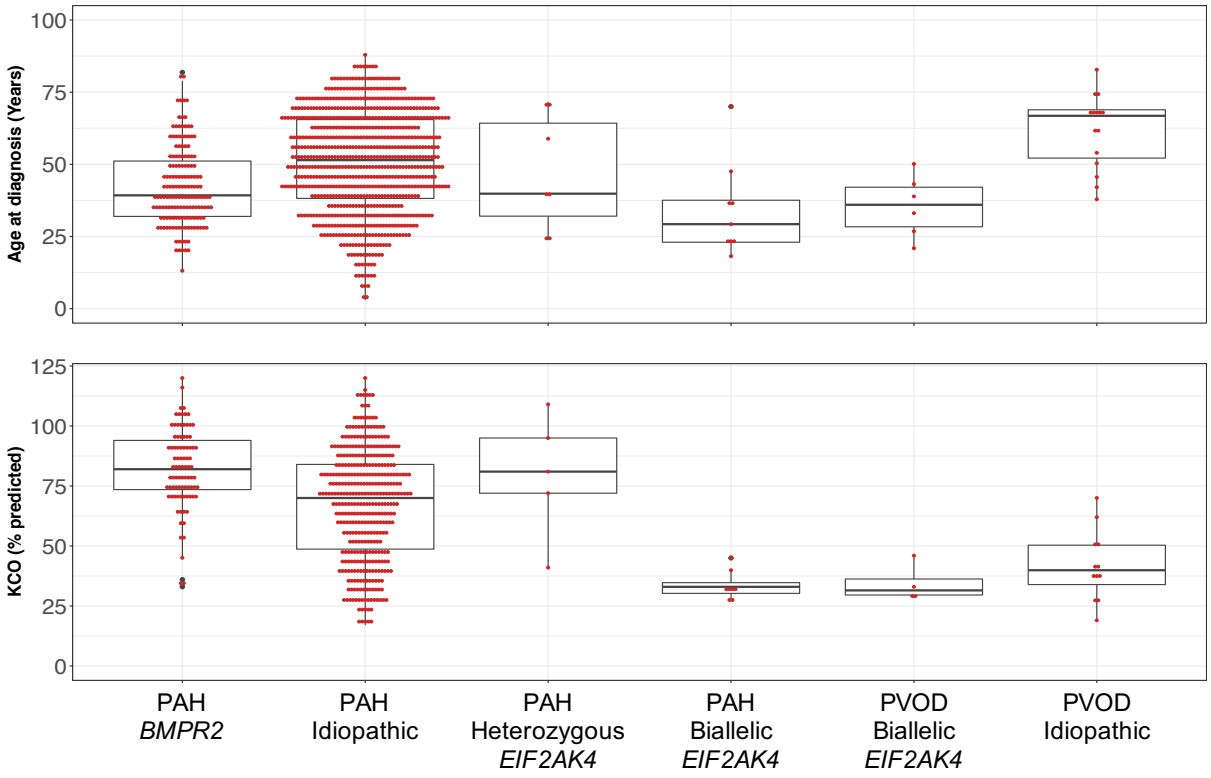


Figure 19. Distributions of age at diagnosis and KCO % predicted in the NIHR BRIDGE PAH Study

Patients recruited to the NIHR BRIDGE PAH Study were categorised by genotype and clinical diagnosis (idiopathic PAH [i.e. no rare or predicted deleterious variants identified in the previously identified disease associated genes]; n = 772), PAH *BMPR2* (n = 157), PAH heterozygous *EIF2AK4* (n = 7), PAH biallelic *EIF2AK4* (n = 9), PVOD biallelic *EIF2AK4* (n = 6), idiopathic PVOD (i.e. no rare or predicted deleterious *EIF2AK4* variants identified; n = 15). KCO % predicted not available for all patients. Patients with either a clinical diagnosis of PAH or PVOD / PCH and carrying biallelic *EIF2AK4* variants had a similar phenotype with a young age at diagnosis (29 years [23 – 38] vs. 36 years [28 -42]) and low KCO % predicted (33 % [30 -35] vs. 32 % [30 – 36]). Although, patients with idiopathic PVOD / PCH also had a low KCO % predicted (40 % [34 – 50]) they were diagnosed later in life (67 years [52 – 69]).

CT features of patients with PAH and biallelic *EIF2AK4* variants

The presence of centrilobular ground glass opacification, mediastinal lymphadenopathy and interlobular septal thickening on CT scans of the chest are suggestive of PVOD / PCH. To investigate whether or not patients with a clinical diagnosis of PAH and biallelic *EIF2AK4* variants were misdiagnosed as PAH, CT scans of a subgroup of patients with a clinical diagnosis of idiopathic PAH (n = 21), PAH and *BMPR2* variants (n = 21), PAH and biallelic

EIF2AK4 variants (n = 7) and patients with a clinical diagnosis of PVOD / PCH (n = 14) were assessed (Table 20, *Dr Nicholas Screaton and Dr Andy Swift*).

Subtle or gross centrilobular ground glass opacification was found in 38 % of patients with idiopathic PAH and 67 % of PAH patients with *BMPR2* variants. This was not significantly different compared to patients with a clinical diagnosis of PAH and biallelic *EIF2AK4* variants (86 %) and patients with a clinical diagnosis of PVOD / PCH (50 %). Gross interlobular septal thickening and mediastinal lymphadenopathy was significantly more frequent amongst patients with PAH and biallelic *EIF2AK4* variants (29 % and 57 % respectively) and those with PVOD / PCH (64 % and 79 %) compared to patients with idiopathic PAH (5 % and 0 %) or PAH and *BMPR2* variants (5 % and 10 %).

Despite the presence of these radiological features in patients with PAH and biallelic *EIF2AK4* variants a radiological suspicion of PVOD / PCH was only raised in 57 %. This compared to a radiological suspicion of PVOD / PCH in 71 % of patients with a clinical diagnosis of PVOD / PCH, 14 % of patients with a clinical diagnosis of idiopathic PAH and 5 % of those with PAH and *BMPR2* variants. Consequently, radiological appearances alone cannot accurately be used to distinguish between patients with a clinical diagnosis of PAH and biallelic *EIF2AK4* variants and patients with idiopathic PAH.

Table 20. CT features of patients with PAH by genotype

		PAH *BMPR2*	Idiopathic PAH	PAH biallelic *EIF2AK4*	PVOD / PCH	p
n		21	21	7	14	
Centrilobular ground glass opacification density	**None**	7 [33.3%]	13 [61.9%]	1 [14.3%]	7 [50.0%]	0.122
	Subtle	12 [57.1%]	5 [23.8%]	2 [28.6%]	3 [21.4%]	
	Present	2 [9.5%]	3 [14.3%]	4 [57.1%]	4 [28.6%]	
Centrilobular ground glass opacification extent	**None**	8 [38.1%]	13 [61.9%]	1 [4.3%]	8 [57.1%]	0.077
	<5%	0 [0.0%]	3 [14.3%]	1 [14.3%]	1 [7.1%]	
	5-25%	2 [9.5%]	0 [0.0%]	2 [28.6%]	1 [7.1%]	
	25-50%	2 [9.5%]	4 [19.0%]	0 [0.0%]	2 [14.3%]	
	50-75%	5 [23.8%]	1 [4.8%]	2 [28.6%]	0 [0.0%]	
	75-100%	4 [19.0%]	0 [0.0%]	1 [14.3%]	2 [14.3%]	
Interlobular septal thickening	**None**	17 [81.0%]	18 [85.7%]	5 [71.4%]	4 [28.6%]	0.001
	Subtle	3 [14.3%]	2 [9.5%]	0 [0.0%]	1 [7.1%]	
	Present	1 [4.8%]	1 [4.8%]	2 [28.6%]	9 [64.3%]	

Mediastinal lymphadenopathy	**None**	19 [90.5%]	21 [100.0%]	3 [42.9%]	3 [21.4%]	<0.001
	Present	2 [9.5%]	0 [0.0%]	4 [57.1%]	11 [78.6%]	
Pleural effusion	**None**	17 [81.0%]	21 [100.0%]	7 [100.0%]	10 [71.4%]	0.048
	Small	4 [19.0%]	0 [0.0%]	9 [0.0%]	4 [28.6%]	
Neovascularity	**None**	12 [57.1%]	18 [85.7%]	6 [85.7%]	13 [92.9%]	0.077
	Present	9 [42.9%]	3 [14.3%]	1 [14.3%]	1 [7.1%]	
CT diagnosis	**PAH**	20 [95.2%]	18 [85.7%]	3 [42.9%]	4 [28.6%]	
	PVOD / PCH	1 [4.8%]	3 [4.3%]	4 [57.1%]	10 [71.4%]	

Histological features of a patient with PAH and a homozygous *EIF2AK4* variant

The gold standard for the diagnosis of PVOD / PCH remains histological assessment of the lung parenchyma. However, lung biopsy can be associated with significant morbidity in patients with elevated pulmonary artery pressures. Therefore, lung tissue is often only assessed after lung transplantation or post mortem.

From the NIHR BRIDGE PAH Study, just one patient with a clinical diagnosis of idiopathic PAH and carrying a biallelic *EIF2AK4* variant underwent lung transplantation and had lung tissue available for assessment. The patient was diagnosed at 22 years of age with idiopathic PAH. He had no family history of the disease. At presentation he had severe pulmonary hypertension with significant right ventricular impairment (mPAP 65 mmHg, PVR 18 WU, CI 1.4 L/min/m^2) and was in functional class 3. His KCO was 31 % predicted despite normal spirometry. Review of his HRCT chest revealed only subtle centrilobular groundglass opacification, no mediastinal lymphadenopathy and no interlobular septal thickening. There was no suspicion of PVOD / PCH on review of his HRCT chest.

This patient had a homozygous *EIF2AK4* missense variant (c.1795G>C, p.Gly599Arg). The missense variant was not reported in the ExAC database, occurred in a conserved area of the genome (GERP score 5.5) and was predicted to be deleterious (CADD score 32, PolyPhen-2 prediction “probably damaging [1]”, SIFT prediction “deleterious [0]”). Interestingly, the same homozygous variant was also identified in a second unrelated patient with a clinical diagnosis of idiopathic PAH.

The patient underwent lung transplantation and the explanted lung tissue examined. The predominant histological feature was pulmonary arterial vasculopathy (*Dr Peter Dorfmüller*; Figure 20). The pulmonary arteries predominantly showed concentric and eccentric intimal fibrosis . No plexiform lesions were observed. Although infrequent, there was some fibrosis of the septal veins and venules, some of which were nearly completely occluded. Although there was evidence of capillary congestion, no capillary hemangiomatosis was observed. The histological features, predominantly pulmonary arteriopathy, supported the clinical and radiological diagnosis of idiopathic PAH.

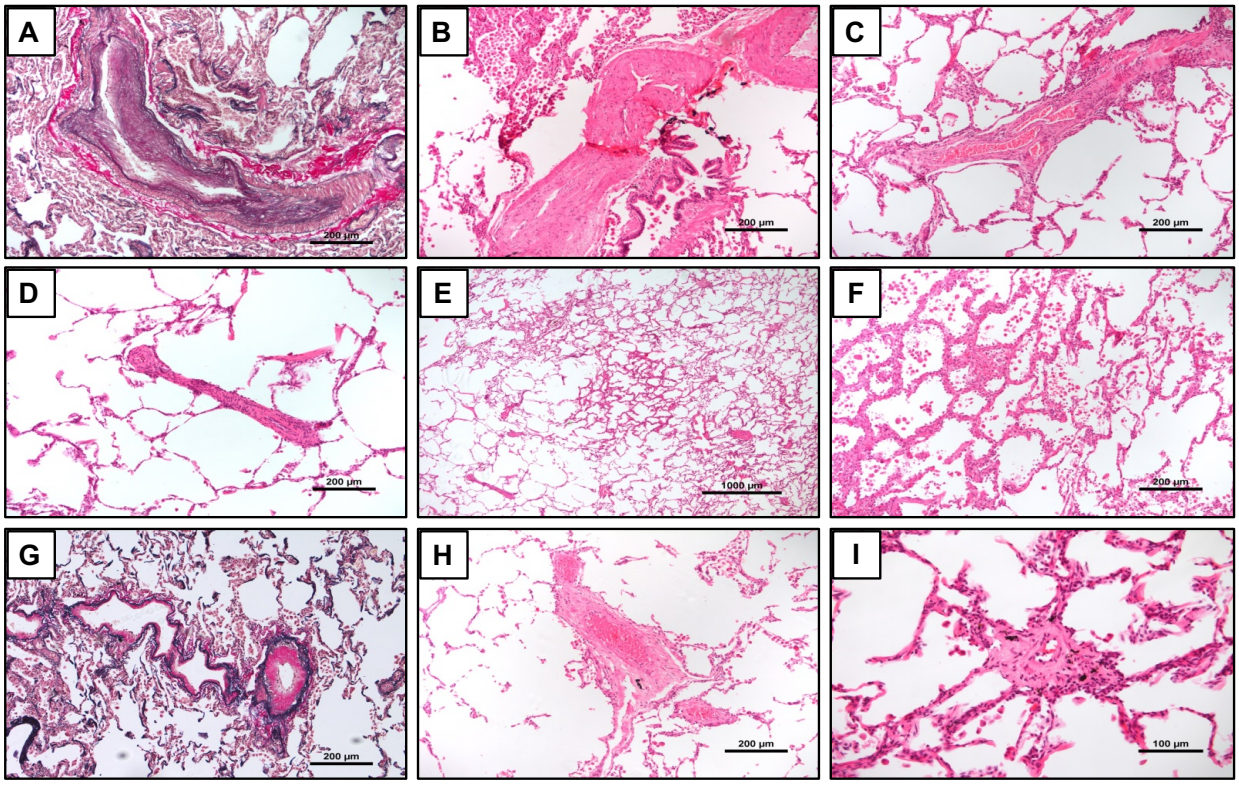


Figure 20. Representative histopathological images from a patient with a clinical diagnosis of idiopathic PAH and a rare and predicted deleterious homozygous *EIF2AK4* variant

The pulmonary arteries showed eccentric and concentric intimal fibrosis and medial hypertrophy (A, B) as well as some lesions with features of recanalised thrombus (C). Several concentrically muscularised arterioles were also observed (D). No complex plexiform lesions were present. There was patchy thickening of the alveolar septa with capillary congestion and pigmented intra-alveolar macrophages similar to PCH (E, F). Venous remodelling was difficult to trace and infrequent, but present. Fibrous thickening of the intima in septal veins (G, I) and a micro-vessel (H). [Photomicrographs by Dr Dorfmüller].

Response to treatment in patients with PAH and biallelic *EIF2AK4* variants

As patients with PVOD / PCH have a worse prognosis compared to patients with idiopathic PAH, the clinical outcomes of patients with biallelic *EIF2AK4* variants and a clinical diagnosis of PAH were assessed. The response to pulmonary artery vasodilator therapies analysis was restricted to idiopathic PAH patients and PAH patients with *BMPR2* variants included in the radiological analysis described above as well as all PAH patients with biallelic *EIF2AK4* variants.

The response to pulmonary artery vasodilator therapies after 1 and 3 years was assessed (Table 21). The change in functional class at both 1 and 3 years was significantly worse in patients with a clinical diagnosis of PAH and biallelic *EIF2AK4* variants compared to patients with idiopathic PAH and PAH patients with *BMPR2* variants. PAH patients with biallelic *EIF2AK4* variants showed no change in functional class whereas patients with idiopathic PAH or PAH patients with *BMPR2* variants improved their functional class by 1 class. The proportion of patients receiving prostanoids therapies was similar between the three groups. None of the PAH patients with biallelic *EIF2AK4* variants developed pulmonary oedema.

Table 21. Response to pulmonary artery vasodilator therapies

Group	Time to assessment 1 (days)	n	Change in 6mwt distance (m)	Change in FC	Time to assessment 2 (days)	n	Change in 6mwt distance (m)	Change in FC	Number on prostanoid therapy before the 2nd assessment [%]
PAH *BMPR2*	357 [314 - 386]	21	+69 [20 - 100]	-1 [-1 - -1]	1120 [1055 - 1174]	18	+45 [31 - 115]	-1 [-1 - -0.5]	5 [23%]
PAH biallelic *EIF2AK4*	358 [335 -388]	9	+28 [-13 - 77]	0 [-1 - 0]	1102 [1090 - 1112]	5	+62 [-8 - 132]	0 [0 - 0]	1 [10%]
Idiopathic PAH	387 [340 - 414]	16	+81 [61 - 151]	-1 [-1 - 0]	1118 [1105 - 1159]	9	+104 [20 - 144]	-1 [-1 - 0]	4 [17%]
p	0.295		0.343	0.039	0.730		0.748	0.044	0.816

6mwt – six-minute walk test, FC – functional class. Drop in number of patients between assessment 1 and 2 due to death, transplantation or lack of sufficient follow up time. Data presented as median [IQR] unless stated. Kruskal-Wallis test used to assess differences between groups.

Prognosis of patients with PAH and biallelic *EIF2AK4* variants

A Cox proportional hazards analysis was performed to assess whether or not patients with a clinical diagnosis of PAH and biallelic *EIF2AK4* variants had a worse prognosis compared to other patients with PAH. Eight-hundred and seventy-four patients with survival data recorded were included in the analysis. Patients with variants in genes other than *BMPR2* or biallelic *EIF2AK4* variants were excluded. Age at diagnosis, gender and whether patients were incident / prevalent cases were included as covariates.

In an untruncated survival analysis PAH patients with biallelic *EIF2AK4* variants had a significantly higher hazard ratio compared to patients with idiopathic PAH (5.00 [2.02 - 12.36]; $p < 0.001$). Older age (1.04 [1.03 - 1.05]; $p < 0.001$), male gender (1.57 [1.18 - 2.09]; $p = 0.002$) and incident cases (2.11 [1.40 - 3.17]; $p < 0.001$) were associated with significantly worse survival in this model. The model met the proportional hazards assumption and assessment of the Cox-Snell residuals suggested a good model fit. A left truncated survival analysis including age, gender and incident / prevalent status as covariates did not have a good model fit, likely due to small numbers. Figure 21 shows a Kaplan-Meier plot demonstrating the effect of genotype on survival.

In a multivariate Cox proportional hazards model including KCO % predicted as a covariate, patients with a clinical diagnosis of PAH and biallelic *EIF2AK4* variants did not have a significantly worse prognosis compared to patients with idiopathic PAH (2.34 [0.87 - 6.33]; $p = 0.094$).

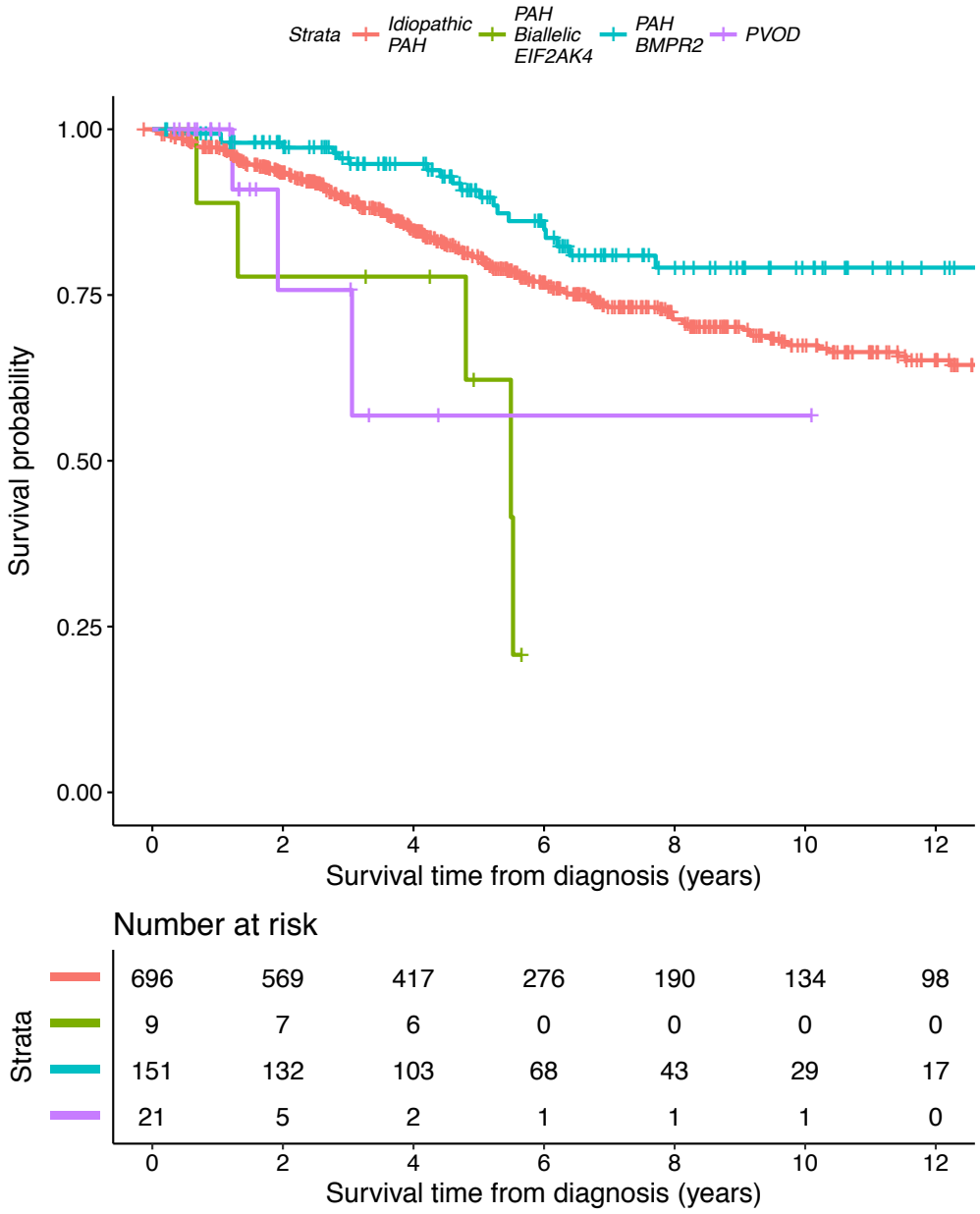


Figure 21. Kaplan-Meier plot showing differences in survival by genotype

The time from diagnosis to death was calculated where survival data was available. Patients were censored at the time of last contact or transplantation. Patients were categorised based on clinical diagnosis and genotype: idiopathic PAH (n = 696), PAH *BMPR2* variant carriers (n = 151), PAH biallelic *EIF2AK4* variant carriers (n = 9) and idiopathic PVOD (n = 21). In an untruncated multivariate Cox proportional hazards model (including age, gender, incident/prevalent status, and *EIF2AK4* genotype) PAH patients with biallelic *EIF2AK4* variants had a higher hazards ratio (5.00 [2.02 - 12.36]; $p < 0.001$) compared to patients with idiopathic PAH.

Biallelic *EIF2AK4* variant carrier phenotype

The above analyses demonstrate that patients with biallelic *EIF2AK4* variants share many clinical and radiological characteristics regardless of their clinical diagnosis (either PVOD / PCH or PAH). However, the spectrum of clinical, radiological and histological features associated with PVOD / PCH is broader than has been previously recognised. Consequently, the presence of only subtle or not readily apparent features in patients with biallelic *EIF2AK4* variants may result in a clinical diagnosis of PAH.

Assuming a molecular diagnosis of PVOD / PCH in all patients with biallelic *EIF2AK4* variants, differences between patients with idiopathic PVOD / PCH and patients with biallelic *EIF2AK4* variants were reassessed. Patients with biallelic *EIF2AK4* variants were younger at diagnosis (33 years [IQR: 24 - 41]) compared to those with idiopathic PVOD / PCH (67 years [52 - 69], p = 0.001). No significant difference was seen in KCO between the two groups (33 % predicted [30 - 35] and 40 % predicted [34 - 50]; p = 0.557). However, in the CT subgroup analysis centrilobular ground glass opacification appeared more extensive in those with biallelic *EIF2AK4* variants (82 %) compared to those with idiopathic PVOD / PCH (10 %; p = 0.012). Furthermore, pleural effusions were more common amongst those with idiopathic PVOD / PCH (40 %) compared to patients with biallelic *EIF2AK4* variants (0 %, p = 0.035). This may suggest that patients with biallelic *EIF2AK4* variants have a distinct radiological phenotype compared to patients with idiopathic PVOD / PCH.

Discussion

This is the largest and most comprehensive description of patients with a clinical diagnosis of PAH carrying biallelic *EIF2AK4* variants. Previously the presence of biallelic *EIF2AK4* variants were reported in patients with a clear clinical diagnosis of PVOD / PCH as well as a large kindred and a single family with a possible diagnosis of PAH (20-22, 93, 306). Assuming that all patients with biallelic *EIF2AK4* variants have a unified molecular diagnosis of PVOD / PCH, novel radiographic differences were identified between patients with idiopathic PVOD / PCH and those carrying biallelic *EIF2AK4* variants.

The discovery of biallelic *EIF2AK4* variants in PVOD / PCH raised the possibility of a rapid molecular diagnosis in the majority of patients with familial, and up to 25 % of patients with sporadic PVOD / PCH (20, 21). In this study, the presence of biallelic *EIF2AK4* variants was associated with a poor prognosis, even in patients who had a clinical diagnosis of PAH, and who did not develop pulmonary oedema in response to pulmonary artery vasodilator therapies. Therefore, early identification of these patients through clinical genetic testing may prompt early referral for lung transplantation similar to patients with clinically diagnosed PVOD / PCH (314).

The presence of biallelic *EIF2AK4* variants in patients with a clinical diagnosis of PAH raises the question whether *EIF2AK4* variants can cause classical idiopathic PAH, or whether there are cases of PVOD / PCH caused by biallelic *EIF2AK4* variants that are wrongly classified even by expert PH centres. This study shows that current clinical, radiological and histological assessments can be difficult to interpret. The presence of subtle or infrequent features may lead to an incorrect diagnosis of PAH in patients with biallelic *EIF2AK4* variants. Consequently, it is important to recognise that patients with pathogenic biallelic *EIF2AK4* variants may present with a spectrum of phenotypic, radiological and histological features that can overlap with PAH.

PAH patients with biallelic *EIF2AK4* variants demonstrated a reduced K_{CO} despite normal spirometry, which is characteristic of patients with PVOD / PCH. The reduced K_{CO} likely reflects widespread reduction in alveolar gas exchange due to endothelial proliferation and

patchy thickening of the blood gas barrier by the process of capillary haemangiomatosis. Ultrastructural thickening of the capillary basement membrane may also play a role (372). In keeping with previous reports in PVOD / PCH, PAH patients with biallelic variants in *EIF2AK4* were younger at diagnosis than patients with either *BMPR2* variants or patients with idiopathic PAH (20, 306). However, the presence of these characteristic features has a low positive predictive value for the identification of patients with biallelic *EIF2AK4* variants.

In contrast to previous descriptions of patients with PVOD / PCH, none of the patients with clinically diagnosed PAH and biallelic *EIF2AK4* variants developed pulmonary oedema in response to pulmonary artery vasodilator therapies. In classical PVOD / PCH patients, pulmonary oedema with intravenous prostanoids has been reported in up to 44% of patients after a median treatment duration of just 9 days (300). Presumably the extent and severity of the pulmonary venous involvement in these patients might underlie the differing responses to pulmonary artery vasodilators.

It is generally considered that HRCT imaging is a useful non-invasive test to assist in the diagnosis of suspected PVOD / PCH (281). Although there was an increased prevalence of mediastinal lymphadenopathy and interlobular septal thickening in PAH patients with biallelic *EIF2AK4* variants, the radiological features at the time of diagnosis could not accurately determine the underlying molecular diagnosis (222). The differing radiological features of all patients with biallelic *EIF2AK4* variants compared to patients with idiopathic PVOD / PCH is of interest and a novel finding. This may reflect differences between the younger onset genetic cases of PVOD / PCH, compared with the predominantly older group of patients without *EIF2AK4* variants in whom other non-genetic factors, such as exposure to inorganic solvents, may play an important role (294). Similar findings have been reported by Montani et al. They describe a younger age at disease onset for patients carrying biallelic *EIF2AK4* variants, but no other clinical or prognostic differences were identified (306). Their study did not assess radiological features.

Histological examination (usually post mortem or from explanted lungs) is often considered essential for diagnostic confirmation of PVOD / PCH but may be confounded by the heterogeneous nature of vascular pathology (109). Surgical biopsy of the lung in patients with

severe PAH is contraindicated. A limitation of this study is that lung tissue from only one patient with biallelic *EIF2AK4* variants was available for analysis. This patient had a rare and predicted deleterious homozygous missense variant in *EIF2AK4*. The predominant feature on assessment of the explanted lung tissue was of pulmonary arteriopathy, as usually seen in PAH. Although only infrequent, fibrosis of the septal venules and the possible presence of siderophages in the alveolar space were observed. These features are found in patients with PVOD / PCH. This case supports the hypothesis that patients with biallelic *EIF2AK4* variants may present with a spectrum of pulmonary venous and arterial involvement.

There are increasing reports of phenotypic, radiological and histological similarities between PAH and PVOD / PCH (112, 222, 296). Tenorio et al. reported a homozygous missense variant in *EIF2AK4* in a large kindred of Iberian Romani with apparent heritable PAH (22). This kindred is likely to have PVOD / PCH as these diagnoses were not confirmed histologically and PVOD was suspected in half the patients. More recently, Best et al. also report two sisters with apparent heritable PAH carrying biallelic *EIF2AK4* variants (93). These patients also had a reduced KCO but had not had HRCT assessment of their lung parenchyma which may have altered their clinical diagnosis. Taken together, these previous reports are compatible with the findings in this larger cohort, that patients with a clinical presentation of idiopathic or heritable PAH may in fact have underlying PVOD / PCH as determined by genetic analysis.

A strength of this study is the centralised reporting of radiographic features. However, the data collection was retrospective and incomplete in some cases. Assessing rare diseases, such as PAH and PVOD / PCH, with a prospective study recruiting incident cases would take a prohibitively long time. This is especially true when assessing survival and response to therapy. In this study, that included prevalent and retrospectively recruited patients, a worse prognosis in patients with a clinical diagnosis of PAH and biallelic *EIF2AK4* variants was demonstrated. However, the inclusion of prevalent and retrospectively recruited patients can introduce biases such as immortal time bias, when there are extended periods between diagnosis and enrolment in the study, and survivor bias. The effect of these biases and confounders can be difficult to predict. In all groups there are likely to be patients who died prior to study enrolment, and thus would not feature in any analysis. A sensitivity (left truncated) analysis, attempting to eliminate these sources of bias, by including only

prospectively recruited patients from the UK, did not have sufficient power to show a difference in survival between genotypes. Larger studies of survival and response to therapy will be needed to definitively show whether “misclassified” PAH patients with biallelic *EIF2AK4* variants have a similarly poor prognosis as classical PVOD / PCH patients.

In previous studies variants in both *EIF2AK4* alleles are required to cause PVOD / PCH (20, 21). In autosomal recessive disorders, it is unusual for the heterozygous state to manifest the disease phenotype and thus heterozygous *EIF2AK4* variants would not be expected to be pathogenic. The finding of heterozygous *EIF2AK4* variants is of interest in light of a recent report suggesting that they may alter disease penetrance in patients with *BMPR2* variants (371). Further studies are required to determine whether heterozygous *EIF2AK4* variants contribute to disease pathogenesis in PAH.

Summary

The study identified a high proportion of biallelic *EIF2AK4* variants in patients with a clinical diagnosis of PVOD / PCH. In addition, 9 patients with a clinical diagnosis of PAH were also shown to carry biallelic *EIF2AK4* variants. These patients had a characteristic early age at diagnosis and low KCO, similar to that previously described in heritable PVOD / PCH. In addition, these patients had a poor prognosis compared to patients with idiopathic PAH despite the fact none developed pulmonary oedema in response to pulmonary artery vasodilator therapies.

The range of clinical, radiological and histological findings associated with biallelic *EIF2AK4* variants is broader than previously recognised. Therefore, the presence of subtle or infrequent changes previously associated with PVOD / PCH may result in a misclassification of these patients as PAH, although the true molecular diagnosis is PVOD / PCH. Ascertaining the *EIF2AK4* variant status of patients through clinical genetic testing provides additional information to aid risk stratification and guide management. In a young patient presenting with apparent PAH, the presence of a low KCO with normal spirometry strongly suggests the presence of underlying biallelic *EIF2AK4* variants.

Discussion

The clinical classification for pulmonary hypertension has evolved in parallel with our understanding of the disease. The current clinical classification, as agreed by consensus at the 5th World Symposium on Pulmonary Hypertension in 2013, groups the disease based on similarities in disease pathogenesis, haemodynamics, associated conditions, treatment options and more recently the underlying molecular diagnosis (1). Our knowledge of the genetic basis of PAH has advanced rapidly over recent years. Previously, rare variants in 10 genes have been associated with PAH or PVOD / PCH. More recently, 4 further novel genes (*SOX17*, *AQP1*, *ATP13A3* and *GDF2*) have been implicated in disease pathogenesis through the NIHR BRIDGE PAH Study (373). However, only 24 % of patients recruited to the study were identified as carrying rare or predicted deleterious variants in genes associated with the disease. The majority of patients in the study had no causal agent (environmental or genetic) identified and were thus classified as idiopathic PAH – a diagnosis of exclusion. Consequently, the presenting clinical features, response to treatment and prognoses of patients with idiopathic PAH is very varied. Such heterogeneity hinders accurate risk stratification and a more personalised treatment approach, as well as complicating the interpretation of randomised clinical trials. Furthermore, much work is required to ascertain the clinical consequences of carrying rare and predicted deleterious variants in disease associated genes.

In this Thesis, I have sought to identify clinically important subgroups of patients with idiopathic PAH as well as identify novel phenotype-genotype associations in patients carrying variants in disease associated genes. The availability of WGS data for all patients was pivotal to this as it allowed the accurate and unbiased categorisation of patients based on their genetic profile. While the breadth of phenotype data collected from the time of diagnosis allowed detailed characterisation of these patients.

Comorbidities and utility of Kco measurement

Amongst patients with idiopathic PAH, an important finding was the identification of mild left ventricular diastolic dysfunction and small left to right cardiac shunts leading to physiological changes that may have played a role in disease pathogenesis. The changing demographics of patients with idiopathic PAH has been noted previously. Compared to the first NIH registry,

published in 1987, patients recruited to this study and other contemporaneous registries are much older (26, 177). It may be hypothesised that this reflects the increasing prevalence of obesity and other cardiovascular risk factors in developed countries. These risk factors can result in subtle left ventricular diastolic dysfunction despite preserved systolic function. Unfortunately, the classification of these patients can be difficult if they develop pulmonary hypertension but their PCWP remains < 15 mmHg at rest when measured by right heart catheterisation (374). By current haemodynamic definitions these patients would have PAH and more specifically idiopathic PAH. The current clinical guidelines state that a combination of clinical risk factors and clinical assessments should be used to identify patients with Group 2 pulmonary hypertension. However, as demonstrated, many patients with these risk factors and left atrial dilation are classified by expert centres as idiopathic PAH.

I have shown that older male patients have the worst prognosis amongst patients with idiopathic PAH. This group of patients are more likely to have cardiovascular risk factors and left ventricular diastolic dysfunction (indicated by left atrial dilation). However, the presence of cardiovascular risk factors or left atrial dilation were not of prognostic significance. This may be due to incomplete assessment of cardiovascular risk factors leading to misclassification of patients and a lack of study power. Additional clinical assessments are required to fully assess the significance of left ventricular diastolic dysfunction in patients with PAH. For example, assessment of pulmonary artery compliance or detailed cardiac MRI studies may yield further insights in to disease pathogenesis and aid further classification of patients (375, 376).

An important observation was the fact that KCO was also reduced in older patients with cardiovascular comorbidities and normal spirometry. This reduction in KCO may simply reflect increased pulmonary venous pressure or may suggest that there is pulmonary venous remodelling similar to that seen in PVOD / PCH. A recent study demonstrated that the pulmonary circulation remodelling seen in patients with heart failure with preserved ejection fraction (HFpEF) was similar to that seen in PVOD. These patients also had a reduced D_LCO (377). Consequently, it could be hypothesised that the low KCO observed in these older patients with idiopathic PAH reflects pulmonary venous remodelling. If validated a low KCO could be used as an accurate and simple non-invasive marker to identify patients with

clinically important diastolic left ventricular dysfunction analogous to its use in identifying patients with PVOD / PCH or biallelic *EIF2AK4* variants. The association between venous remodelling and a low KCO may also explain the prognostic significance of a low KCO observed in patients with idiopathic PAH.

Differential remodelling of the pulmonary arterial and venous circulation has been used to differentiate between PAH and PVOD / PCH. However, in this Thesis it has been demonstrated that the spectrum of histological changes associated with PVOD / PCH is greater than previously thought and has significant overlap with PAH. Although the explanted lungs of just a single patient with biallelic *EIF2AK4* variants and a clinical diagnosis of idiopathic PAH was available for assessment, CT scans were assessed in larger numbers of patients. It was shown that the radiographic features associated with PVOD / PCH (and pulmonary venous remodelling) were not always present (or only subtle) in patients with biallelic *EIF2AK4* variants and a clinical diagnosis of idiopathic PAH. However, all patients with biallelic *EIF2AK4* variants had a characteristic reduction in KCO. These observations also highlight the importance of assessing KCO in patients with pulmonary hypertension.

The range of clinical, radiological and histological features associated with biallelic *EIF2AK4* variants results in difficulties in identifying patients carrying these variants based solely on current clinical assessments. Clinical genetic testing may help identify these patients earlier, allowing accurate risk stratification, earlier referral for transplant assessment and enable genetic counselling for at risk family members.

Another important group of patients identified in this Thesis were the young patients (< 50 years at diagnosis) with normal spirometry and a low KCO (< 50 % predicted). Unlike the older patients with a low KCO, these patients did not have evidence of left heart disease and were found not to have any obvious genetic cause for the disease (including common variation in *EIF2AK4* or variants in other genes in the integrated stress response pathway [data not shown]). Their low KCO suggests that they too may have pulmonary venopathy. Further investigation of the cause of their disease and their prognosis would be important. In particular whether or not they have a poor outcome similar to patients with biallelic *EIF2AK4* variants should be explored.

Variation in *BMPR2* and other disease associated genes

Variants in *BMPR2* account for most patients with heritable PAH. In this Thesis I have been able to confirm that patients with variants in *BMPR2* are younger at diagnosis and have more severe pulmonary haemodynamic impairment. However, the prognosis of these patients did not appear to be worse than patients with idiopathic PAH, unlike previously reported by Evans et al. in a much larger study (30). Despite its size, this study may still lack statistical power to demonstrate any difference in survival.

Novel associations with blood count indices were also observed. Characterisation of the effects of loss of *BMPR2* signalling is an important step towards developing treatments aimed at restoring this signalling pathway. For example, exogenous BMP9 (the ligand of BMPR2) has been shown to reverse and prevent pulmonary hypertension in preclinical models (152). Understanding the full clinical consequences of pathogenic *BMPR2* variation may be of value in monitoring the response to these novel treatments.

The majority of patients with PAH do not appear to have a rare and deleterious variant causing their disease. However, common genetic variation in combination with environmental and other genetic risk factors may account for additional cases. Previous studies have used a candidate gene approach to investigate common genetic variation in PAH (99, 122, 170, 178). The only published GWAS in PAH to date has identified common variation in *CBLN2* as a potential risk factor for the development of PAH (102). The NIHR BRIDGE PAH Study and other international studies now provide an opportunity to reassess common variation in PAH. A publication looking at the role of common variation arising in this study is currently in submission. Further study of the phenotypes of patients carrying these common variants would be of value to understand their clinical impact.

Future perspectives

There are many opportunities arising from the NIHR BRIDGE PAH Study and the MRC / BHF Cohort Study. In addition to the specific questions raised earlier in the discussion, a simple but important analysis will be to revisit the baseline diagnostic data after a period of time. Re-running the analyses described in this Thesis on a more mature dataset, with more

individuals and more events in the survival analyses, may yield further insights. However, this is unlikely to significantly increase the number of patients with rare variants in genes other than *BMPR2*. Efforts are already underway to combine the NIHR BRIDGE PAH Study WGS data with a whole exome sequencing study (National Biological Sample and Data Repository for Pulmonary Arterial Hypertension, also known as PAH Biobank) from the USA. Such large international collaborations will be required to identify further novel genetic variation involved in disease pathogenesis and to elucidate specific genotype-phenotype associations for most genes.

In this Thesis I have used simple comparative statistics and rank regression models to identify genotype-phenotype associations. Alternative statistical methodologies may be employed in the future to identify specific subgroups of patients. Cluster analyses have been used in other diseases to identify clinically important subgroups (378, 379). However, meaningful application of these methods requires the selection of variables that may reflect disease pathogenesis, severity or outcomes. As individuals with incomplete datasets are excluded, this has hindered application of these methods to the current study. However, the results presented in this Thesis may be used to guide selection of variables for capture in future studies that will utilise these statistical methods.

In addition to baseline diagnostic data, the MRC / BHF Cohort Study is collecting longitudinal clinical data from patients. This will provide a valuable dataset with which to explore response to pulmonary artery vasodilator therapies in specific groups of patients. Currently, there is no information available to suggest that one class of pulmonary artery vasodilator therapy is superior to another in specific groups or individuals.

Publications

Identification of rare sequence variation underlying heritable pulmonary arterial hypertension. Gräf S, Haimel M, Bleda M, Hadinnapola C* et al. Nature Communications. 2018 Apr 12;9(1):1416 *joint first author

Phenotypic characterisation of *EIF2AK4* mutation carriers in a large cohort of patients diagnosed clinically with pulmonary arterial hypertension. Hadinnapola C, Bleda M, Haimel M et al. Circulation. 2017 Nov 21;136(21):2022-2033.

"Plasma metabolomics implicates modified transfer RNAs and altered bioenergetics in the outcomes of pulmonary arterial hypertension." Rhodes C, Ghataorhe P, Wharton J, Rue-Albrecht K, Hadinnapola C et al. Circulation. 2017 Jan 31;135(5):460-475.

References

1. Galie N, Humbert M, Vachiery JL, Gibbs S, Lang I, Torbicki A, et al. 2015 ESC/ERS Guidelines for the diagnosis and treatment of pulmonary hypertension: The Joint Task Force for the Diagnosis and Treatment of Pulmonary Hypertension of the European Society of Cardiology (ESC) and the European Respiratory Society (ERS): Endorsed by: Association for European Paediatric and Congenital Cardiology (AEPC), International Society for Heart and Lung Transplantation (ISHLT). Eur Heart J. 2016;37(1):67-119.
2. Hatano S, Strasser T, Organization WH. Primary pulmonary hypertension : report on a WHO meeting, Geneva, 15-17 October 1973 Hatano S, Strasser T, editors. Geneva: World Health Organization 1975. 45 p.
3. Simonneau G, Galie N, Rubin LJ, Langleben D, Seeger W, Domenighetti G, et al. Clinical classification of pulmonary hypertension. J Am Coll Cardiol. 2004;43(12 Suppl S):5s-12s.
4. Simonneau G, Gatzoulis MA, Adatia I, Celermajer D, Denton C, Ghofrani A, et al. Updated clinical classification of pulmonary hypertension. J Am Coll Cardiol. 2013;62(25 Suppl):D34-41.
5. Abenhaim L, Moride Y, Brenot F, Rich S, Benichou J, Kurz X, et al. Appetite-suppressant drugs and the risk of primary pulmonary hypertension. International Primary Pulmonary Hypertension Study Group. N Engl J Med. 1996;335(9):609-16.
6. Montani D, Seferian A, Savale L, Simonneau G, Humbert M. Drug-induced pulmonary arterial hypertension: a recent outbreak. Eur Respir Rev. 2013;22(129):244-50.
7. Chin KM, Channick RN, Rubin LJ. Is methamphetamine use associated with idiopathic pulmonary arterial hypertension? Chest. 2006;130(6):1657-63.
8. Thomson JR, Machado RD, Pauciulo MW, Morgan NV, Humbert M, Elliott GC, et al. Sporadic primary pulmonary hypertension is associated with germline mutations of the gene encoding BMPR-II, a receptor member of the TGF-beta family. J Med Genet. 2000;37(10):741-5.
9. Koehler R, Grunig E, Pauciulo MW, Hoeper MM, Olschewski H, Wilkens H, et al. Low frequency of BMPR2 mutations in a German cohort of patients with sporadic idiopathic pulmonary arterial hypertension. J Med Genet. 2004;41(12):e127.

10. Morisaki H, Nakanishi N, Kyotani S, Takashima A, Tomoike H, Morisaki T. *BMPR2* mutations found in Japanese patients with familial and sporadic primary pulmonary hypertension. Hum Mutat. 2004;23(6):632.

11. Harrison RE, Berger R, Haworth SG, Tulloh R, Mache CJ, Morrell NW, et al. Transforming growth factor-beta receptor mutations and pulmonary arterial hypertension in childhood. Circulation. 2005;111(4):435-41.

12. Sankelo M, Flanagan JA, Machado R, Harrison R, Rudarakanchana N, Morrell N, et al. BMPR2 mutations have short lifetime expectancy in primary pulmonary hypertension. Hum Mutat. 2005;26(2):119-24.

13. Fujiwara M, Yagi H, Matsuoka R, Akimoto K, Furutani M, Imamura S, et al. Implications of mutations of activin receptor-like kinase 1 gene (*ALK1*) in addition to bone morphogenetic protein receptor II gene (BMPR2) in children with pulmonary arterial hypertension. Circ J. 2008;72(1):127-33.

14. Pfarr N, Szamalek-Hoegel J, Fischer C, Hinderhofer K, Nagel C, Ehlken N, et al. Hemodynamic and clinical onset in patients with hereditary pulmonary arterial hypertension and BMPR2 mutations. Respir Res. 2011;12:99.

15. Liu D, Liu QQ, Eyries M, Wu WH, Yuan P, Zhang R, et al. Molecular genetics and clinical features of Chinese idiopathic and heritable pulmonary arterial hypertension patients. Eur Respir J. 2012;39(3):597-603.

16. Kabata H, Satoh T, Kataoka M, Tamura Y, Ono T, Yamamoto M, et al. Bone morphogenetic protein receptor type 2 mutations, clinical phenotypes and outcomes of Japanese patients with sporadic or familial pulmonary hypertension. Respirology. 2013;18(7):1076-82.

17. Machado RD, Pauciulo MW, Thomson JR, Lane KB, Morgan NV, Wheeler L, et al. BMPR2 haploinsufficiency as the inherited molecular mechanism for primary pulmonary hypertension. Am J Hum Genet. 2001;68(1):92-102.

18. Lane KB, Machado RD, Pauciulo MW, Thomson JR, Phillips JA, 3rd, Loyd JE, et al. Heterozygous germline mutations in *BMPR2*, encoding a TGF-beta receptor, cause familial primary pulmonary hypertension. Nat Genet. 2000;26(1):81-4.

19. Deng Z, Morse JH, Slager SL, Cuervo N, Moore KJ, Venetos G, et al. Familial primary pulmonary hypertension (gene pph1) is caused by mutations in the bone morphogenetic protein receptor–II gene. Am J Hum Genet. 2000;67(3):737-44.

20. Eyries M, Montani D, Girerd B, Perret C, Leroy A, Lonjou C, et al. *EIF2AK4* mutations cause pulmonary veno-occlusive disease, a recessive form of pulmonary hypertension. Nat Genet. 2014;46(1):65-9.

21. Best DH, Sumner KL, Austin ED, Chung WK, Brown LM, Borczuk AC, et al. *EIF2AK4* mutations in pulmonary capillary hemangiomatosis. Chest. 2014;145(2):231-6.

22. Tenorio J, Navas P, Barrios E, Fernandez L, Nevado J, Quezada CA, et al. A founder *EIF2AK4* mutation causes an aggressive form of pulmonary arterial hypertension in Iberian Gypsies. Clin Genet. 2015;88(6):579-83.

23. Union E. Medicines for rare diseases - orphan drugs: European Union; 2011 [Available from: http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=LEGISSUM:l21167.

24. Haffner ME. Adopting orphan drugs--two dozen years of treating rare diseases. N Engl J Med. 2006;354(5):445-7.

25. Ling Y, Johnson MK, Kiely DG, Condliffe R, Elliot CA, Gibbs JS, et al. Changing demographics, epidemiology, and survival of incident pulmonary arterial hypertension: results from the pulmonary hypertension registry of the United Kingdom and Ireland. Am J Respir Crit Care Med. 2012;186(8):790-6.

26. Frost AE, Badesch DB, Barst RJ, Benza RL, Elliott CG, Farber HW, et al. The changing picture of patients with pulmonary arterial hypertension in the United States: how REVEAL differs from historic and non-US Contemporary Registries. Chest. 2011;139(1):128-37.

27. Humbert M, Sitbon O, Chaouat A, Bertocchi M, Habib G, Gressin V, et al. Pulmonary arterial hypertension in France: results from a national registry. Am J Respir Crit Care Med. 2006;173(9):1023-30.

28. Centre HaSCI. National Audit of Pulmonary Hypertension 2015: Health and Social Care Information Centre; 2016 [Available from: http://www.content.digital.nhs.uk/catalogue/PUB20043/nati-pulm-hype-audi-2015-rep.pdf.

29. Barbara G, David M, Xavier J, Mélanie E, Azzedine Y, Benjamin S, et al. Genetic counselling in a national referral centre for pulmonary hypertension. 2016.

30. Evans JD, Girerd B, Montani D, Wang XJ, Galie N, Austin ED, et al. *BMPR2* mutations and survival in pulmonary arterial hypertension: an individual participant data meta-analysis. Lancet Respir Med. 2016;4(2):129-37.

31. Melmon KL, Braunwald E. Familial pulmonary hypertension. N Engl J Med. 1963;269:770-5.

32. Dresdale DT, Michtom RJ, Schultz M. Recent studies in primary pulmonary hypertension, including pharmacodynamic observations on pulmonary vascular resistance. Bull N Y Acad Med. 1954;30(3):195-207.
33. Loyd JE, Primm RK, Newman JH. Familial primary pulmonary hypertension: clinical patterns. Am Rev Respir Dis. 1984;129(1):194-7.
34. Hood WB, Jr., Spencer H, Lass RW, Daley R. Primary pulmonary hypertension: familial occurrence. Br Heart J. 1968;30(3):336-43.
35. Nichols WC, Koller DL, Slovis B, Foroud T, Terry VH, Arnold ND, et al. Localization of the gene for familial primary pulmonary hypertension to chromosome 2q31-32. Nat Genet. 1997;15(3):277-80.
36. Morse JH, Jones AC, Barst RJ, Hodge SE, Wilhelmsen KC, Nygaard TG. Mapping of familial primary pulmonary hypertension locus (PPH1) to chromosome 2q31-q32. Circulation. 1997;95(12):2603-6.
37. Pulst SM. Genetic linkage analysis. Arch Neurol. 1999;56(6):667-72.
38. Cogan J, Austin E, Hedges L, Womack B, West J, Loyd J, et al. Role of BMPR2 alternative splicing in heritable pulmonary arterial hypertension penetrance. Circulation. 2012;126(15):1907-16.
39. Cogan JD, Vnencak-Jones CL, Phillips JA, 3rd, Lane KB, Wheeler LA, Robbins IM, et al. Gross BMPR2 gene rearrangements constitute a new cause for primary pulmonary hypertension. Genet Med. 2005;7(3):169-74.
40. Cogan JD, Pauciulo MW, Batchman AP, Prince MA, Robbins IM, Hedges LK, et al. High frequency of *BMPR2* exonic deletions/duplications in familial pulmonary arterial hypertension. Am J Respir Crit Care Med. 2006;174(5):590-8.
41. Aldred MA, Vijayakrishnan J, James V, Soubrier F, Gomez-Sanchez MA, Martensson G, et al. BMPR2 gene rearrangements account for a significant proportion of mutations in familial and idiopathic pulmonary arterial hypertension. Hum Mutat. 2006;27(2):212-3.
42. Machado RD, Southgate L, Eichstaedt CA, Aldred MA, Austin ED, Best DH, et al. Pulmonary Arterial Hypertension: A Current Perspective on Established and Emerging Molecular Genetic Defects. Hum Mutat. 2015;36(12):1113-27.
43. Viales RR, Eichstaedt CA, Ehlken N, Fischer C, Lichtblau M, Grunig E, et al. Mutation in *BMPR2* Promoter: A 'Second Hit' for Manifestation of Pulmonary Arterial Hypertension? PLoS One. 2015;10(7):e0133042.

44. Wang H, Li W, Zhang W, Sun K, Song X, Gao S, et al. Novel promoter and exon mutations of the *BMPR2* gene in Chinese patients with pulmonary arterial hypertension. Eur J Hum Genet. 2009;17(8):1063-9.
45. Pousada G, Lupo V, Castro-Sanchez S, Alvarez-Satta M, Sanchez-Monteagudo A, Baloira A, et al. Molecular and functional characterization of the *BMPR2* gene in pulmonary arterial hypertension. Sci Rep. 2017;7(1):1923.
46. Aldred MA, Machado RD, James V, Morrell NW, Trembath RC. Characterization of the *BMPR2* 5'-untranslated region and a novel mutation in pulmonary hypertension. Am J Respir Crit Care Med. 2007;176(8):819-24.
47. Machado RD, James V, Southwood M, Harrison RE, Atkinson C, Stewart S, et al. Investigation of second genetic hits at the *BMPR2* locus as a modulator of disease progression in familial pulmonary arterial hypertension. Circulation. 2005;111(5):607-13.
48. Larkin EK, Newman JH, Austin ED, Hemnes AR, Wheeler L, Robbins IM, et al. Longitudinal analysis casts doubt on the presence of genetic anticipation in heritable pulmonary arterial hypertension. Am J Respir Crit Care Med. 2012;186(9):892-6.
49. Hamid R, Cogan JD, Hedges LK, Austin E, Phillips JA, 3rd, Newman JH, et al. Penetrance of pulmonary arterial hypertension is modulated by the expression of normal *BMPR2* allele. Hum Mutat. 2009;30(4):649-54.
50. Atkinson C, Stewart S, Upton PD, Machado R, Thomson JR, Trembath RC, et al. Primary pulmonary hypertension is associated with reduced pulmonary vascular expression of type II bone morphogenetic protein receptor. Circulation. 2002;105(14):1672-8.
51. Dewachter L, Adnot S, Guignabert C, Tu L, Marcos E, Fadel E, et al. Bone morphogenetic protein signalling in heritable versus idiopathic pulmonary hypertension. Eur Respir J. 2009;34(5):1100-10.
52. Aldred MA, Comhair SA, Varella-Garcia M, Asosingh K, Xu W, Noon GP, et al. Somatic chromosome abnormalities in the lungs of patients with pulmonary arterial hypertension. Am J Respir Crit Care Med. 2010;182(9):1153-60.
53. Kawabata M, Chytil A, Moses HL. Cloning of a novel type II serine/threonine kinase receptor through interaction with the type I transforming growth factor-beta receptor. J Biol Chem. 1995;270(10):5625-30.

54. Liu F, Ventura F, Doody J, Massague J. Human type II receptor for bone morphogenic proteins (BMPs): extension of the two-kinase receptor model to the BMPs. Mol Cell Biol. 1995;15(7):3479-86.

55. Wong WK, Knowles JA, Morse JH. Bone morphogenetic protein receptor type II C-terminus interacts with c-Src: implication for a role in pulmonary arterial hypertension. Am J Respir Cell Mol Biol. 2005;33(5):438-46.

56. Johnson JA, Hemnes AR, Perrien DS, Schuster M, Robinson LJ, Gladson S, et al. Cytoskeletal defects in *BMPR2*-associated pulmonary arterial hypertension. Am J Physiol Lung Cell Mol Physiol. 2012;302(5):L474-84.

57. Machado RD, Rudarakanchana N, Atkinson C, Flanagan JA, Harrison R, Morrell NW, et al. Functional interaction between BMPR-II and Tctex-1, a light chain of Dynein, is isoform-specific and disrupted by mutations underlying primary pulmonary hypertension. Hum Mol Genet. 2003;12(24):3277-86.

58. Foletta VC, Lim MA, Soosairajah J, Kelly AP, Stanley EG, Shannon M, et al. Direct signaling by the BMP type II receptor via the cytoskeletal regulator LIMK1. J Cell Biol. 2003;162(6):1089-98.

59. Nishihara A, Watabe T, Imamura T, Miyazono K. Functional heterogeneity of bone morphogenetic protein receptor-II mutants found in patients with primary pulmonary hypertension. Mol Biol Cell. 2002;13(9):3055-63.

60. David L, Mallet C, Mazerbourg S, Feige JJ, Bailly S. Identification of BMP9 and BMP10 as functional activators of the orphan activin receptor-like kinase 1 (ALK1) in endothelial cells. Blood. 2007;109(5):1953-61.

61. Lebrin F, Goumans MJ, Jonker L, Carvalho RL, Valdimarsdottir G, Thorikay M, et al. Endoglin promotes endothelial cell proliferation and TGF-beta/ALK1 signal transduction. Embo j. 2004;23(20):4018-28.

62. Panchenko MP, Williams MC, Brody JS, Yu Q. Type I receptor serine-threonine kinase preferentially expressed in pulmonary blood vessels. Am J Physiol. 1996;270(4 Pt 1):L547-58.

63. Johnson DW, Berg JN, Baldwin MA, Gallione CJ, Marondel I, Yoon SJ, et al. Mutations in the activin receptor-like kinase 1 gene in hereditary haemorrhagic telangiectasia type 2. Nat Genet. 1996;13(2):189-95.

64. Vorselaars V, Velthuis S, van Gent M, Westermann C, Snijder R, Mager J, et al. Pulmonary hypertension in a large cohort with hereditary hemorrhagic telangiectasia. Respiration. 2017.
65. Lyle MA, Fenstad ER, McGoon MD, Frantz RP, Krowka MJ, Kane GC, et al. Pulmonary hypertension in hereditary hemorrhagic telangiectasia. Chest. 2016;149(2):362-71.
66. Trembath RC, Thomson JR, Machado RD, Morgan NV, Atkinson C, Winship I, et al. Clinical and molecular genetic features of pulmonary hypertension in patients with hereditary hemorrhagic telangiectasia. N Engl J Med. 2001;345(5):325-34.
67. Harrison RE, Flanagan JA, Sankelo M, Abdalla SA, Rowell J, Machado RD, et al. Molecular and functional analysis identifies *ALK-1* as the predominant cause of pulmonary hypertension related to hereditary haemorrhagic telangiectasia. J Med Genet. 2003;40(12):865-71.
68. McAllister KA, Grogg KM, Johnson DW, Gallione CJ, Baldwin MA, Jackson CE, et al. Endoglin, a TGF-beta binding protein of endothelial cells, is the gene for hereditary haemorrhagic telangiectasia type 1. Nat Genet. 1994;8(4):345-51.
69. Chaouat A, Coulet F, Favre C, Simonneau G, Weitzenblum E, Soubrier F, et al. Endoglin germline mutation in a patient with hereditary haemorrhagic telangiectasia and dexfenfluramine associated pulmonary arterial hypertension. Thorax. 2004;59(5):446-8.
70. Nasim MT, Ogo T, Ahmed M, Randall R, Chowdhury HM, Snape KM, et al. Molecular genetic characterization of SMAD signaling molecules in pulmonary arterial hypertension. Hum Mutat. 2011;32(12):1385-9.
71. Shintani M, Yagi H, Nakayama T, Saji T, Matsuoka R. A new nonsense mutation of SMAD8 associated with pulmonary arterial hypertension. J Med Genet. 2009;46(5):331-7.
72. Chida A, Shintani M, Nakayama T, Furutani Y, Hayama E, Inai K, et al. Missense mutations of the *BMPR1B* (ALK6) gene in childhood idiopathic pulmonary arterial hypertension. Circ J. 2012;76(6):1501-8.
73. Wang G, Fan R, Ji R, Zou W, Penny DJ, Varghese NP, et al. Novel homozygous *BMP9* nonsense mutation causes pulmonary arterial hypertension: a case report. BMC Pulm Med. 2016;16:17.
74. Wooderchak-Donahue WL, McDonald J, O'Fallon B, Upton PD, Li W, Roman BL, et al. BMP9 mutations cause a vascular-anomaly syndrome with phenotypic overlap with hereditary hemorrhagic telangiectasia. Am J Hum Genet. 2013;93(3):530-7.

75. Ma L, Roman-Campos D, Austin ED, Eyries M, Sampson KS, Soubrier F, et al. A novel channelopathy in pulmonary arterial hypertension. N Engl J Med. 2013;369(4):351-61.
76. Olschewski A, Li Y, Tang B, Hanze J, Eul B, Bohle RM, et al. Impact of TASK-1 in human pulmonary artery smooth muscle cells. Circ Res. 2006;98(8):1072-80.
77. Navas Tejedor P, Tenorio Castano J, Palomino Doza J, Arias Lajara P, Gordo Trujillo G, Lopez Meseguer M, et al. An homozygous mutation in KCNK3 is associated with an aggressive form of hereditary pulmonary arterial hypertension. Clin Genet. 2017;91(3):453-7.
78. Higasa K, Ogawa A, Terao C, Shimizu M, Kosugi S, Yamada R, et al. A burden of rare variants in *BMPR2* and *KCNK3* contributes to a risk of familial pulmonary arterial hypertension. BMC Pulm Med. 2017;17(1):57.
79. Gilbert G, Ducret T, Savineau JP, Marthan R, Quignard JF. Caveolae are involved in mechanotransduction during pulmonary hypertension. Am J Physiol Lung Cell Mol Physiol. 2016;310(11):L1078-87.
80. Austin ED, Ma L, LeDuc C, Berman Rosenzweig E, Borczuk A, Phillips JA, 3rd, et al. Whole exome sequencing to identify a novel gene (caveolin-1) associated with human pulmonary arterial hypertension. Circ Cardiovasc Genet. 2012;5(3):336-43.
81. Schrauwen I, Szelinger S, Siniard AL, Kurdoglu A, Corneveaux JJ, Malenica I, et al. A Frame-shift mutation in *CAV1* is associated with a severe neonatal progeroid and lipodystrophy syndrome. PLoS One. 2015;10(7):e0131797.
82. Garg A, Kircher M, Del Campo M, Amato RS, Agarwal AK. Whole exome sequencing identifies de novo heterozygous *CAV1* mutations associated with a novel neonatal onset lipodystrophy syndrome. Am J Med Genet A. 2015;167a(8):1796-806.
83. Han B, Copeland CA, Kawano Y, Rosenzweig EB, Austin ED, Shahmirzadi L, et al. Characterization of a caveolin-1 mutation associated with both pulmonary arterial hypertension and congenital generalized lipodystrophy. Traffic. 2016;17(12):1297-312.
84. Cao H, Alston L, Ruschman J, Hegele RA. Heterozygous *CAV1* frameshift mutations (MIM 601047) in patients with atypical partial lipodystrophy and hypertriglyceridemia. Lipids Health Dis. 2008;7:3.
85. Kim CA, Delepine M, Boutet E, El Mourabit H, Le Lay S, Meier M, et al. Association of a homozygous nonsense caveolin-1 mutation with Berardinelli-Seip congenital lipodystrophy. J Clin Endocrinol Metab. 2008;93(4):1129-34.

86. Gomez J, Reguero JR, Alvarez C, Junquera MR, Arango A, Moris C, et al. A semiconductor chip-based next generation sequencing procedure for the main pulmonary hypertension genes. Lung. 2015;193(4):571-4.
87. Ballif BC, Theisen A, Rosenfeld JA, Traylor RN, Gastier-Foster J, Thrush DL, et al. Identification of a recurrent microdeletion at 17q23.1q23.2 flanked by segmental duplications associated with heart defects and limb abnormalities. Am J Hum Genet. 2010;86(3):454-61.
88. Nimmakayalu M, Major H, Sheffield V, Solomon DH, Smith RJ, Patil SR, et al. Microdeletion of 17q22q23.2 encompassing *TBX2* and *TBX4* in a patient with congenital microcephaly, thyroid duct cyst, sensorineural hearing loss, and pulmonary hypertension. Am J Med Genet A. 2011;155a(2):418-23.
89. Kerstjens-Frederikse WS, Bongers EM, Roofthooft MT, Leter EM, Douwes JM, Van Dijk A, et al. *TBX4* mutations (small patella syndrome) are associated with childhood-onset pulmonary arterial hypertension. J Med Genet. 2013;50(8):500-6.
90. Bongers EM, Duijf PH, van Beersum SE, Schoots J, Van Kampen A, Burckhardt A, et al. Mutations in the human *TBX4* gene cause small patella syndrome. Am J Hum Genet. 2004;74(6):1239-48.
91. Levy M, Eyries M, Szezepanski I, Ladouceur M, Nadaud S, Bonnet D, et al. Genetic analyses in a cohort of children with pulmonary hypertension. Eur Respir J. 2016;48(4):1118-26.
92. Navas P, Tenorio J, Quezada CA, Barrios E, Gordo G, Arias P, et al. Molecular analysis of *BMPR2, TBX4, and KCNK3* and genotype-phenotype correlations in Spanish patients and families with idiopathic and hereditary pulmonary arterial hypertension. Rev Esp Cardiol (Engl Ed). 2016;69(11):1011-9.
93. Best DH, Sumner KL, Smith BP, Damjanovich-Colmenares K, Nakayama I, Brown LM, et al. *EIF2AK4* mutations in patients diagnosed with pulmonary arterial hypertension. Chest. 2016.
94. Eichstaedt CA, Song J, Viales RR, Pan Z, Benjamin N, Fischer C, et al. First identification of Kruppel-like factor 2 mutation in heritable pulmonary arterial hypertension. Clin Sci (Lond). 2017;131(8):689-98.

95. Dekker RJ, van Soest S, Fontijn RD, Salamanca S, de Groot PG, VanBavel E, et al. Prolonged fluid shear stress induces a distinct set of endothelial cell genes, most specifically lung Kruppel-like factor (*KLF2*). Blood. 2002;100(5):1689-98.
96. Herrera M, Garvin JL. Novel role of AQP-1 in NO-dependent vasorelaxation. Am J Physiol Renal Physiol. 2007;292(5):F1443-51.
97. Clipson A, Wang M, de Leval L, Ashton-Key M, Wotherspoon A, Vassiliou G, et al. KLF2 mutation is the most frequent somatic change in splenic marginal zone lymphoma and identifies a subset with distinct genotype. Leukemia. 2015;29(5):1177-85.
98. Yuan XJ, Wang J, Juhaszova M, Gaine SP, Rubin LJ. Attenuated K+ channel gene transcription in primary pulmonary hypertension. Lancet. 1998;351(9104):726-7.
99. Remillard CV, Tigno DD, Platoshyn O, Burg ED, Brevnova EE, Conger D, et al. Function of Kv1.5 channels and genetic variations of *KCNA5* in patients with idiopathic pulmonary arterial hypertension. Am J Physiol Cell Physiol. 2007;292(5):C1837-53.
100. Pousada G, Baloira A, Vilarino C, Cifrian JM, Valverde D. Novel mutations in *BMPR2*, *ACVRL1* and *KCNA5* genes and hemodynamic parameters in patients with pulmonary arterial hypertension. PLoS One. 2014;9(6):e100261.
101. Wang G, Knight L, Ji R, Lawrence P, Kanaan U, Li L, et al. Early onset severe pulmonary arterial hypertension with 'two-hit' digenic mutations in both *BMPR2* and *KCNA5* genes. Int J Cardiol. 2014;177(3):e167-9.
102. Germain M, Eyries M, Montani D, Poirier O, Girerd B, Dorfmuller P, et al. Genome-wide association analysis identifies a susceptibility locus for pulmonary arterial hypertension. Nat Genet. 2013;45(5):518-21.
103. de Jesus Perez VA, Yuan K, Lyuksyutova MA, Dewey F, Orcholski ME, Shuffle EM, et al. Whole-exome sequencing reveals *TopBP1* as a novel gene in idiopathic pulmonary arterial hypertension. Am J Respir Crit Care Med. 2014;189(10):1260-72.
104. Dresdale D, Schultz M, Mitchom R. Primary pulmonary hypertension: I. Clinical and hemodynamic study - ScienceDirect. American Journal of Medicine. 1951;11(6):686-94.
105. Wagenvoort CA, Wagenvoort N. Primary pulmonary hypertension: a pathologic study of the lung vessels in 156 clinically diagnosed cases. Circulation. 1970;42(6):1163-84.
106. Stacher E, Graham BB, Hunt JM, Gandjeva A, Groshong SD, McLaughlin VV, et al. Modern age pathology of pulmonary arterial hypertension. Am J Respir Crit Care Med. 2012;186(3):261-72.

107. Moser KM, Bloor CM. Pulmonary vascular lesions occurring in patients with chronic major vessel thromboembolic pulmonary hypertension. Chest. 1993;103(3):685-92.
108. Yi ES, Kim H, Ahn H, Strother J, Morris T, Masliah E, et al. Distribution of obstructive intimal lesions and their cellular phenotypes in chronic pulmonary hypertension. A morphometric and immunohistochemical study. Am J Respir Crit Care Med. 2000;162(4 Pt 1):1577-86.
109. Pietra GG, Capron F, Stewart S, Leone O, Humbert M, Robbins IM, et al. Pathologic assessment of vasculopathies in pulmonary hypertension. J Am Coll Cardiol. 2004;43(12 Suppl S):25s-32s.
110. Heath D, Edwards J. The pathology of hypertensive pulmonary vascular disease. Circulation. 1958;18(4):533-47.
111. Tuder RM. Pulmonary vascular remodeling in pulmonary hypertension. Cell Tissue Res. 2017;367(3):643-9.
112. Ghigna MR, Guignabert C, Montani D, Girerd B, Jais X, Savale L, et al. *BMPR2* mutation status influences bronchial vascular changes in pulmonary arterial hypertension. Eur Respir J. 2016;48(6):1668-81.
113. Tuder RM, Groves B, Badesch DB, Voelkel NF. Exuberant endothelial cell growth and elements of inflammation are present in plexiform lesions of pulmonary hypertension. Am J Pathol. 1994;144(2):275-85.
114. Tamosiuniene R, Tian W, Dhillon G, Wang L, Sung YK, Gera L, et al. Regulatory T cells limit vascular endothelial injury and prevent pulmonary hypertension. Circ Res. 2011;109(8):867-79.
115. Soon E, Holmes AM, Treacy CM, Doughty NJ, Southgate L, Machado RD, et al. Elevated levels of inflammatory cytokines predict survival in idiopathic and familial pulmonary arterial hypertension. Circulation. 2010;122(9):920-7.
116. Humbert M, Monti G, Brenot F, Sitbon O, Portier A, Grangeot-Keros L, et al. Increased interleukin-1 and interleukin-6 serum concentrations in severe primary pulmonary hypertension. Am J Respir Crit Care Med. 1995;151(5):1628-31.
117. Itoh T, Nagaya N, Ishibashi-Ueda H, Kyotani S, Oya H, Sakamaki F, et al. Increased plasma monocyte chemoattractant protein-1 level in idiopathic pulmonary arterial hypertension. Respirology. 2006;11(2):158-63.

118. Sanchez O, Marcos E, Perros F, Fadel E, Tu L, Humbert M, et al. Role of endothelium-derived CC chemokine ligand 2 in idiopathic pulmonary arterial hypertension. Am J Respir Crit Care Med. 2007;176(10):1041-7.

119. Cracowski JL, Chabot F, Labarere J, Faure P, Degano B, Schwebel C, et al. Proinflammatory cytokine levels are linked to death in pulmonary arterial hypertension. Eur Respir J. 2014;43(3):915-7.

120. Hagen M, Fagan K, Steudel W, Carr M, Lane K, Rodman DM, et al. Interaction of interleukin-6 and the BMP pathway in pulmonary smooth muscle. Am J Physiol Lung Cell Mol Physiol. 2007;292(6):L1473-9.

121. Selimovic N, Bergh CH, Andersson B, Sakiniene E, Carlsten H, Rundqvist B. Growth factors and interleukin-6 across the lung circulation in pulmonary hypertension. Eur Respir J. 2009;34(3):662-8.

122. Chaouat A, Savale L, Chouaid C, Tu L, Sztrymf B, Canuet M, et al. Role for interleukin-6 in COPD-related pulmonary hypertension. Chest. 2009;136(3):678-87.

123. Brock M, Trenkmann M, Gay RE, Michel BA, Gay S, Fischler M, et al. Interleukin-6 modulates the expression of the bone morphogenic protein receptor type II through a novel STAT3-microRNA cluster 17/92 pathway. Circ Res. 2009;104(10):1184-91.

124. Song Y, Jones JE, Beppu H, Keaney JF, Jr., Loscalzo J, Zhang YY. Increased susceptibility to pulmonary hypertension in heterozygous *BMPR2*-mutant mice. Circulation. 2005;112(4):553-62.

125. Beppu H, Ichinose F, Kawai N, Jones RC, Yu PB, Zapol WM, et al. *BMPR-II* heterozygous mice have mild pulmonary hypertension and an impaired pulmonary vascular remodeling response to prolonged hypoxia. Am J Physiol Lung Cell Mol Physiol. 2004;287(6):L1241-7.

126. Hurst LA, Dunmore BJ, Long L, Crosby A, Al-Lamki R, Deighton J, et al. TNFalpha drives pulmonary arterial hypertension by suppressing the BMP type-II receptor and altering NOTCH signalling. Nat Commun. 2017;8:14079.

127. Soon E, Crosby A, Southwood M, Yang P, Tajsic T, Toshner M, et al. Bone morphogenetic protein receptor type II deficiency and increased inflammatory cytokine production. A gateway to pulmonary arterial hypertension. Am J Respir Crit Care Med. 2015;192(7):859-72.

128. Chu JW, Kao PN, Faul JL, Doyle RL. High prevalence of autoimmune thyroid disease in pulmonary arterial hypertension. Chest. 2002;122(5):1668-73.
129. Richter MJ, Sommer N, Schermuly R, Grimminger B, Seeger W, Tello K, et al. The prognostic impact of thyroid function in pulmonary hypertension. J Heart Lung Transplant. 2016;35(12):1427-34.
130. Badesch DB, Wynne KM, Bonvallet S, Voelkel NF, Ridgway C, Groves BM. Hypothyroidism and primary pulmonary hypertension: an autoimmune pathogenetic link? Ann Intern Med. 1993;119(1):44-6.
131. Masri FA, Xu W, Comhair SA, Asosingh K, Koo M, Vasanji A, et al. Hyperproliferative apoptosis-resistant endothelial cells in idiopathic pulmonary arterial hypertension. Am J Physiol Lung Cell Mol Physiol. 2007;293(3):L548-54.
132. Yuan JX, Aldinger AM, Juhaszova M, Wang J, Conte JV, Jr., Gaine SP, et al. Dysfunctional voltage-gated K+ channels in pulmonary artery smooth muscle cells of patients with primary pulmonary hypertension. Circulation. 1998;98(14):1400-6.
133. Bonnet S, Michelakis ED, Porter CJ, Andrade-Navarro MA, Thebaud B, Haromy A, et al. An abnormal mitochondrial-hypoxia inducible factor-1alpha-Kv channel pathway disrupts oxygen sensing and triggers pulmonary arterial hypertension in fawn hooded rats: similarities to human pulmonary arterial hypertension. Circulation. 2006;113(22):2630-41.
134. Platoshyn O, Golovina VA, Bailey CL, Limsuwan A, Krick S, Juhaszova M, et al. Sustained membrane depolarization and pulmonary artery smooth muscle cell proliferation. Am J Physiol Cell Physiol. 2000;279(5):C1540-9.
135. Zhao Y, Peng J, Lu C, Hsin M, Mura M, Wu L, et al. Metabolomic heterogeneity of pulmonary arterial hypertension. PLoS One. 2014;9(2):e88727.
136. McMurtry MS, Bonnet S, Wu X, Dyck JR, Haromy A, Hashimoto K, et al. Dichloroacetate prevents and reverses pulmonary hypertension by inducing pulmonary artery smooth muscle cell apoptosis. Circ Res. 2004;95(8):830-40.
137. Michelakis ED, McMurtry MS, Wu XC, Dyck JR, Moudgil R, Hopkins TA, et al. Dichloroacetate, a metabolic modulator, prevents and reverses chronic hypoxic pulmonary hypertension in rats: role of increased expression and activity of voltage-gated potassium channels. Circulation. 2002;105(2):244-50.
138. Guignabert C, Tu L, Izikki M, Dewachter L, Zadigue P, Humbert M, et al. Dichloroacetate treatment partially regresses established pulmonary hypertension in mice

with SM22alpha-targeted overexpression of the serotonin transporter. Faseb j. 2009;23(12):4135-47.

139. Sutendra G, Dromparis P, Bonnet S, Haromy A, McMurtry MS, Bleackley RC, et al. Pyruvate dehydrogenase inhibition by the inflammatory cytokine TNFalpha contributes to the pathogenesis of pulmonary arterial hypertension. J Mol Med (Berl). 2011;89(8):771-83.

140. Sutendra G, Bonnet S, Rochefort G, Haromy A, Folmes KD, Lopaschuk GD, et al. Fatty acid oxidation and malonyl-CoA decarboxylase in the vascular remodeling of pulmonary hypertension. Sci Transl Med. 2010;2(44):44ra58.

141. Pugh ME, Robbins IM, Rice TW, West J, Newman JH, Hemnes AR. Unrecognized glucose intolerance is common in pulmonary arterial hypertension. J Heart Lung Transplant. 2011;30(8):904-11.

142. Zamanian RT, Hansmann G, Snook S, Lilienfeld D, Rappaport KM, Reaven GM, et al. Insulin resistance in pulmonary arterial hypertension. Eur Respir J. 2009;33(2):318-24.

143. Taraseviciute A, Voelkel NF. Severe pulmonary hypertension in postmenopausal obese women. Eur J Med Res. 2006;11(5):198-202.

144. Hansmann G, Wagner RA, Schellong S, Perez VA, Urashima T, Wang L, et al. Pulmonary arterial hypertension is linked to insulin resistance and reversed by peroxisome proliferator-activated receptor-gamma activation. Circulation. 2007;115(10):1275-84.

145. Hansmann G, de Jesus Perez VA, Alastalo TP, Alvira CM, Guignabert C, Bekker JM, et al. An antiproliferative BMP-2/PPARgamma/apoE axis in human and murine SMCs and its role in pulmonary hypertension. J Clin Invest. 2008;118(5):1846-57.

146. Summer R, Fiack CA, Ikeda Y, Sato K, Dwyer D, Ouchi N, et al. Adiponectin deficiency: a model of pulmonary hypertension associated with pulmonary vascular disease. Am J Physiol Lung Cell Mol Physiol. 2009;297(3):L432-8.

147. Ameshima S, Golpon H, Cool CD, Chan D, Vandivier RW, Gardai SJ, et al. Peroxisome proliferator-activated receptor gamma (PPARgamma) expression is decreased in pulmonary hypertension and affects endothelial cell growth. Circ Res. 2003;92(10):1162-9.

148. Belly MJ, Tiede H, Morty RE, Schulz R, Voswinckel R, Tanislav C, et al. HbA1c in pulmonary arterial hypertension: a marker of prognostic relevance? J Heart Lung Transplant. 2012;31(10):1109-14.

149. Yeager ME, Halley GR, Golpon HA, Voelkel NF, Tuder RM. Microsatellite instability of endothelial cell growth and apoptosis genes within plexiform lesions in primary pulmonary hypertension. Circ Res. 2001;88(1):E2-e11.
150. Teichert-Kuliszewska K, Kutryk MJ, Kuliszewski MA, Karoubi G, Courtman DW, Zucco L, et al. Bone morphogenetic protein receptor-2 signaling promotes pulmonary arterial endothelial cell survival: implications for loss-of-function mutations in the pathogenesis of pulmonary hypertension. Circ Res. 2006;98(2):209-17.
151. Majka S, Hagen M, Blackwell T, Harral J, Johnson JA, Gendron R, et al. Physiologic and molecular consequences of endothelial *BMPR2* mutation. Respir Res. 2011;12:84.
152. Long L, Ormiston ML, Yang X, Southwood M, Graf S, Machado RD, et al. Selective enhancement of endothelial BMPR-II with BMP9 reverses pulmonary arterial hypertension. Nat Med. 2015;21(7):777-85.
153. Diebold I, Hennigs JK, Miyagawa K, Li CG, Nickel NP, Kaschwich M, et al. BMPR2 preserves mitochondrial function and DNA during reoxygenation to promote endothelial cell survival and reverse pulmonary hypertension. Cell Metab. 2015;21(4):596-608.
154. Fessel JP, Hamid R, Wittmann BM, Robinson LJ, Blackwell T, Tada Y, et al. Metabolomic analysis of bone morphogenetic protein receptor type 2 mutations in human pulmonary endothelium reveals widespread metabolic reprogramming. Pulm Circ. 2012;2(2):201-13.
155. Hemnes AR, Brittain EL, Trammell AW, Fessel JP, Austin ED, Penner N, et al. Evidence for right ventricular lipotoxicity in heritable pulmonary arterial hypertension. Am J Respir Crit Care Med. 2014;189(3):325-34.
156. McGoon MD, Vanhoutte PM. Aggregating platelets contract isolated canine pulmonary arteries by releasing 5-hydroxytryptamine. J Clin Invest. 1984;74(3):828-33.
157. Nemecek GM, Coughlin SR, Handley DA, Moskowitz MA. Stimulation of aortic smooth muscle cell mitogenesis by serotonin. Proc Natl Acad Sci U S A. 1986;83(3):674-8.
158. Gurtner HP. Aminorex and pulmonary hypertension. A review. Cor Vasa. 1985;27(2-3):160-71.
159. Gaine SP, Rubin LJ, Kmetzo JJ, Palevsky HI, Traill TA. Recreational use of aminorex and pulmonary hypertension. Chest. 2000;118(5):1496-7.
160. Douglas JG, Munro JF, Kitchin AH, Muir AL, Proudfoot AT. Pulmonary hypertension and fenfluramine. Br Med J (Clin Res Ed). 1981;283(6296):881-3.

161. Eddahibi S, Adnot S. Anorexigen-induced pulmonary hypertension and the serotonin (5-HT) hypothesis: lessons for the future in pathogenesis. Respir Res. 2002;3:9.
162. Souza R, Humbert M, Sztrymf B, Jais X, Yaici A, Le Pavec J, et al. Pulmonary arterial hypertension associated with fenfluramine exposure: report of 109 cases. Eur Respir J. 2008;31(2):343-8.
163. Fishman AP. Aminorex to fen/phen: an epidemic foretold. Circulation. 1999;99(1):156-61.
164. Mielke H, Seiler KU, Stumpf U, Wassermann O. [Relation between serotonin metabolism and pulmonary hypertension in rats following administration of various anorectic drugs]. Z Kardiol. 1973;62(12):1090-8.
165. Eddahibi S, Humbert M, Fadel E, Raffestin B, Darmon M, Capron F, et al. Serotonin transporter overexpression is responsible for pulmonary artery smooth muscle hyperplasia in primary pulmonary hypertension. J Clin Invest. 2001;108(8):1141-50.
166. Eddahibi S, Fabre V, Boni C, Martres MP, Raffestin B, Hamon M, et al. Induction of serotonin transporter by hypoxia in pulmonary vascular smooth muscle cells. Relationship with the mitogenic action of serotonin. Circ Res. 1999;84(3):329-36.
167. Marcos E, Fadel E, Sanchez O, Humbert M, Dartevelle P, Simonneau G, et al. Serotonin-induced smooth muscle hyperplasia in various forms of human pulmonary hypertension. Circ Res. 2004;94(9):1263-70.
168. Rothman RB, Baumann MH. Methamphetamine and idiopathic pulmonary arterial hypertension: role of the serotonin transporter. Chest. 2007;132(4):1412-3.
169. Lee SL, Wang WW, Fanburg BL. Dexfenfluramine as a mitogen signal via the formation of superoxide anion. Faseb j. 2001;15(7):1324-5.
170. Zhang H, Xu M, Xia J, Qin RY. Association between serotonin transporter (*SERT*) gene polymorphism and idiopathic pulmonary arterial hypertension: a meta-analysis and review of the literature. Metabolism. 2013;62(12):1867-75.
171. Willers ED, Newman JH, Loyd JE, Robbins IM, Wheeler LA, Prince MA, et al. Serotonin transporter polymorphisms in familial and idiopathic pulmonary arterial hypertension. Am J Respir Crit Care Med. 2006;173(7):798-802.
172. Machado RD, Koehler R, Glissmeyer E, Veal C, Suntharalingam J, Kim M, et al. Genetic association of the serotonin transporter in pulmonary arterial hypertension. Am J Respir Crit Care Med. 2006;173(7):793-7.

173. Chen PI, Cao A, Miyagawa K, Tojais NF, Hennigs JK, Li CG, et al. Amphetamines promote mitochondrial dysfunction and DNA damage in pulmonary hypertension. JCI Insight. 2017;2(2):e90427.
174. Herve P, Launay JM, Scrobohaci ML, Brenot F, Simonneau G, Petitpretz P, et al. Increased plasma serotonin in primary pulmonary hypertension. Am J Med. 1995;99(3):249-54.
175. Lederer DJ, Horn EM, Rosenzweig EB, Karmally W, Jahnes M, Barst RJ, et al. Plasma serotonin levels are normal in pulmonary arterial hypertension. Pulm Pharmacol Ther. 2008;21(1):112-4.
176. Chambers CD, Hernandez-Diaz S, Van Marter LJ, Werler MM, Louik C, Jones KL, et al. Selective serotonin-reuptake inhibitors and risk of persistent pulmonary hypertension of the newborn. N Engl J Med. 2006;354(6):579-87.
177. Rich S, Dantzker DR, Ayres SM, Bergofsky EH, Brundage BH, Detre KM, et al. Primary pulmonary hypertension. A national prospective study. Ann Intern Med. 1987;107(2):216-23.
178. Austin ED, Cogan JD, West JD, Hedges LK, Hamid R, Dawson EP, et al. Alterations in oestrogen metabolism: implications for higher penetrance of familial pulmonary arterial hypertension in females. Eur Respir J. 2009;34(5):1093-9.
179. Austin ED, Hamid R, Hemnes AR, Loyd JE, Blackwell T, Yu C, et al. BMPR2 expression is suppressed by signaling through the estrogen receptor. Biol Sex Differ. 2012;3(1):6.
180. Mair KM, Yang XD, Long L, White K, Wallace E, Ewart MA, et al. Sex affects bone morphogenetic protein type II receptor signaling in pulmonary artery smooth muscle cells. Am J Respir Crit Care Med. 2015;191(6):693-703.
181. White K, Dempsie Y, Nilsen M, Wright AF, Loughlin L, MacLean MR. The serotonin transporter, gender, and 17beta oestradiol in the development of pulmonary arterial hypertension. Cardiovasc Res. 2011;90(2):373-82.
182. Benza RL, Miller DP, Gomberg-Maitland M, Frantz RP, Foreman AJ, Coffey CS, et al. Predicting survival in pulmonary arterial hypertension: insights from the Registry to Evaluate Early and Long-Term Pulmonary Arterial Hypertension Disease Management (REVEAL). Circulation. 2010;122(2):164-72.

183. Humbert M, Sitbon O, Chaouat A, Bertocchi M, Habib G, Gressin V, et al. Survival in patients with idiopathic, familial, and anorexigen-associated pulmonary arterial hypertension in the modern management era. Circulation. 2010;122(2):156-63.
184. Jacobs W, van de Veerdonk MC, Trip P, de Man F, Heymans MW, Marcus JT, et al. The right ventricle explains sex differences in survival in idiopathic pulmonary arterial hypertension. Chest. 2014;145(6):1230-6.
185. Kawut SM, Al-Naamani N, Agerstrand C, Berman Rosenzweig E, Rowan C, Barst RJ, et al. Determinants of right ventricular ejection fraction in pulmonary arterial hypertension. Chest. 2009;135(3):752-9.
186. Frump AL, Goss KN, Vayl A, Albrecht M, Fisher A, Tursunova R, et al. Estradiol improves right ventricular function in rats with severe angioproliferative pulmonary hypertension: effects of endogenous and exogenous sex hormones. Am J Physiol Lung Cell Mol Physiol. 2015;308(9):L873-90.
187. Wright AF, Ewart MA, Mair K, Nilsen M, Dempsie Y, Loughlin L, et al. Oestrogen receptor alpha in pulmonary hypertension. Cardiovasc Res. 2015;106(2):206-16.
188. Mair KM, Wright AF, Duggan N, Rowlands DJ, Hussey MJ, Roberts S, et al. Sex-dependent influence of endogenous estrogen in pulmonary hypertension. Am J Respir Crit Care Med. 2014;190(4):456-67.
189. Chen X, Austin ED, Talati M, Fessel JP, Farber-Eger EH, Brittain EL, et al. Oestrogen inhibition reverses pulmonary arterial hypertension and associated metabolic defects. Eur Respir J. 2017;50(2).
190. Kawut SM, Archer-Chicko CL, DeMichele A, Fritz JS, Klinger JR, Ky B, et al. Anastrozole in pulmonary arterial hypertension. A randomized, double-blind, placebo-controlled trial. Am J Respir Crit Care Med. 2017;195(3):360-8.
191. Alves JL, Jr., Gavilanes F, Jardim C, Fernandes C, Morinaga LTK, Dias B, et al. Pulmonary arterial hypertension in the southern hemisphere: results from a registry of incident Brazilian cases. Chest. 2015;147(2):495-501.
192. Jing ZC, Xu XQ, Han ZY, Wu Y, Deng KW, Wang H, et al. Registry and survival study in chinese patients with idiopathic and familial pulmonary arterial hypertension. Chest. 2007;132(2):373-9.
193. Thenappan T, Ryan JJ, Archer SL. Evolving epidemiology of pulmonary arterial hypertension. Am J Respir Crit Care Med. 2012;186(8):707-9.

194. Hoeper MM, Huscher D, Ghofrani HA, Delcroix M, Distler O, Schweiger C, et al. Elderly patients diagnosed with idiopathic pulmonary arterial hypertension: results from the COMPERA registry. Int J Cardiol. 2013;168(2):871-80.

195. Kawut SM, Horn EM, Berekashvili KK, Garofano RP, Goldsmith RL, Widlitz AC, et al. New predictors of outcome in idiopathic pulmonary arterial hypertension. Am J Cardiol. 2005;95(2):199-203.

196. Brown LM, Chen H, Halpern S, Taichman D, McGoon MD, Farber HW, et al. Delay in recognition of pulmonary arterial hypertension: factors identified from the REVEAL Registry. Chest. 2011;140(1):19-26.

197. Wilkens H, Grimminger F, Hoeper M, Stahler G, Ehlken B, Plesnila-Frank C, et al. Burden of pulmonary arterial hypertension in Germany. Respir Med. 2010;104(6):902-10.

198. Strange G, Gabbay E, Kermeen F, Williams T, Carrington M, Stewart S, et al. Time from symptoms to definitive diagnosis of idiopathic pulmonary arterial hypertension: The delay study. Pulm Circ. 2013;3(1):89-94.

199. Achouh L, Montani D, Garcia G, Jais X, Hamid AM, Mercier O, et al. Pulmonary arterial hypertension masquerading as severe refractory asthma. Eur Respir J. 2008;32(2):513-6.

200. Le RJ, Fenstad ER, Maradit-Kremers H, McCully RB, Frantz RP, McGoon MD, et al. Syncope in adults with pulmonary arterial hypertension. J Am Coll Cardiol. 2011;58(8):863-7.

201. Colman R, Whittingham H, Tomlinson G, Granton J. Utility of the physical examination in detecting pulmonary hypertension. A mixed methods study. PLoS One. 2014;9(10):e108499.

202. D'Alonzo GE, Barst RJ, Ayres SM, Bergofsky EH, Brundage BH, Detre KM, et al. Survival in patients with primary pulmonary hypertension. Results from a national prospective registry. Ann Intern Med. 1991;115(5):343-9.

203. Nickel N, Golpon H, Greer M, Knudsen L, Olsson K, Westerkamp V, et al. The prognostic impact of follow-up assessments in patients with idiopathic pulmonary arterial hypertension. Eur Respir J. 2012;39(3):589-96.

204. Sztrymf B, Coulet F, Girerd B, Yaici A, Jais X, Sitbon O, et al. Clinical outcomes of pulmonary arterial hypertension in carriers of *BMPR2* mutation. Am J Respir Crit Care Med. 2008;177(12):1377-83.

205. Rich S, Kaufmann E, Levy PS. The effect of high doses of calcium-channel blockers on survival in primary pulmonary hypertension. N Engl J Med. 1992;327(2):76-81.
206. Sitbon O, Humbert M, Jais X, Ioos V, Hamid AM, Provencher S, et al. Long-term response to calcium channel blockers in idiopathic pulmonary arterial hypertension. Circulation. 2005;111(23):3105-11.
207. Sitbon O, Humbert M, Jagot JL, Taravella O, Fartoukh M, Parent F, et al. Inhaled nitric oxide as a screening agent for safely identifying responders to oral calcium-channel blockers in primary pulmonary hypertension. Eur Respir J. 1998;12(2):265-70.
208. Elliott CG, Glissmeyer EW, Havlena GT, Carlquist J, McKinney JT, Rich S, et al. Relationship of BMPR2 mutations to vasoreactivity in pulmonary arterial hypertension. Circulation. 2006;113(21):2509-15.
209. Tongers J, Schwerdtfeger B, Klein G, Kempf T, Schaefer A, Knapp JM, et al. Incidence and clinical relevance of supraventricular tachyarrhythmias in pulmonary hypertension. Am Heart J. 2007;153(1):127-32.
210. Forfia PR, Fisher MR, Mathai SC, Housten-Harris T, Hemnes AR, Borlaug BA, et al. Tricuspid annular displacement predicts survival in pulmonary hypertension. Am J Respir Crit Care Med. 2006;174(9):1034-41.
211. Baggen VJ, Driessen MM, Post MC, van Dijk AP, Roos-Hesselink JW, van den Bosch AE, et al. Echocardiographic findings associated with mortality ortransplant in patients with pulmonary arterial hypertension:A systematic review and meta-analysis. Neth Heart J. 2016;24(6):374-89.
212. Harbaum L, Hennigs JK, Baumann HJ, Luneburg N, Griesch E, Bokemeyer C, et al. N-terminal pro-brain natriuretic peptide is a useful prognostic marker in patients with pre-capillary pulmonary hypertension and renal insufficiency. PLoS One. 2014;9(4):e94263.
213. Carrington M, Murphy NF, Strange G, Peacock A, McMurray JJ, Stewart S. Prognostic impact of pulmonary arterial hypertension: a population-based analysis. Int J Cardiol. 2008;124(2):183-7.
214. Rhodes CJ, Wharton J, Howard LS, Gibbs JS, Wilkins MR. Red cell distribution width outperforms other potential circulating biomarkers in predicting survival in idiopathic pulmonary arterial hypertension. Heart. 2011;97(13):1054-60.

215. Taguchi H, Kataoka M, Yanagisawa R, Kawakami T, Tamura Y, Fukuda K, et al. Platelet level as a new prognostic factor for idiopathic pulmonary arterial hypertension in the era of combination therapy. Circ J. 2012;76(6):1494-500.
216. Shah SJ, Thenappan T, Rich S, Tian L, Archer SL, Gomberg-Maitland M. Association of serum creatinine with abnormal hemodynamics and mortality in pulmonary arterial hypertension. Circulation. 2008;117(19):2475-83.
217. Quarck R, Nawrot T, Meyns B, Delcroix M. C-reactive protein: a new predictor of adverse outcome in pulmonary arterial hypertension. J Am Coll Cardiol. 2009;53(14):1211-8.
218. Rhodes CJ, Howard LS, Busbridge M, Ashby D, Kondili E, Gibbs JS, et al. Iron deficiency and raised hepcidin in idiopathic pulmonary arterial hypertension: clinical prevalence, outcomes, and mechanistic insights. J Am Coll Cardiol. 2011;58(3):300-9.
219. Nagaya N, Uematsu M, Satoh T, Kyotani S, Sakamaki F, Nakanishi N, et al. Serum uric acid levels correlate with the severity and the mortality of primary pulmonary hypertension. Am J Respir Crit Care Med. 1999;160(2):487-92.
220. Larsen CM, McCully RB, Murphy JG, Kushwaha SS, Frantz RP, Kane GC. Usefulness of high-density lipoprotein cholesterol to predict survival in pulmonary arterial hypertension. Am J Cardiol. 2016;118(2):292-7.
221. Sandoval J, Bauerle O, Palomar A, Gomez A, Martinez-Guerra ML, Beltran M, et al. Survival in primary pulmonary hypertension. Validation of a prognostic equation. Circulation. 1994;89(4):1733-44.
222. Rajaram S, Swift AJ, Condliffe R, Johns C, Elliot CA, Hill C, et al. CT features of pulmonary arterial hypertension and its major subtypes: a systematic CT evaluation of 292 patients from the ASPIRE Registry. Thorax. 2015;70(4):382-7.
223. Resten A, Maitre S, Humbert M, Sitbon O, Capron F, Simoneau G, et al. Pulmonary arterial hypertension: thin-section CT predictors of epoprostenol therapy failure. Radiology. 2002;222(3):782-8.
224. van Wolferen SA, Marcus JT, Boonstra A, Marques KM, Bronzwaer JG, Spreeuwenberg MD, et al. Prognostic value of right ventricular mass, volume, and function in idiopathic pulmonary arterial hypertension. Eur Heart J. 2007;28(10):1250-7.
225. Swift AJ, Capener D, Johns C, Hamilton N, Rothman A, Elliot C, et al. Magnetic resonance imaging in the prognostic evaluation of patients with pulmonary arterial hypertension. Am J Respir Crit Care Med. 2017;196(2):228-39.

226. Barst RJ, Rubin LJ, Long WA, McGoon MD, Rich S, Badesch DB, et al. A comparison of continuous intravenous epoprostenol (prostacyclin) with conventional therapy for primary pulmonary hypertension. N Engl J Med. 1996;334(5):296-301.
227. Groepenhoff H, Vonk-Noordegraaf A, van de Veerdonk MC, Boonstra A, Westerhof N, Bogaard HJ. Prognostic relevance of changes in exercise test variables in pulmonary arterial hypertension. PLoS One. 2013;8(9):e72013.
228. Miyamoto S, Nagaya N, Satoh T, Kyotani S, Sakamaki F, Fujita M, et al. Clinical correlates and prognostic significance of six-minute walk test in patients with primary pulmonary hypertension. Comparison with cardiopulmonary exercise testing. Am J Respir Crit Care Med. 2000;161(2 Pt 1):487-92.
229. Gall H, Felix JF, Schneck FK, Milger K, Sommer N, Voswinckel R, et al. The Giessen Pulmonary Hypertension Registry: Survival in pulmonary hypertension subgroups. J Heart Lung Transplant. 2017;36(9):957-67.
230. Leter EM, Boonstra AB, Postma FB, Gille JJ, Meijers-Heijboer EJ, Vonk Noordegraaf A. Genetic counselling for pulmonary arterial hypertension: a matter of variable variability. Neth Heart J. 2011;19(2):89-92.
231. Frydman N, Steffann J, Girerd B, Frydman R, Munnich A, Simonneau G, et al. Pre-implantation genetic diagnosis in pulmonary arterial hypertension due to BMPR2 mutation. Eur Respir J. 2012;39(6):1534-5.
232. Chung WK, Austin ED, Best DH, Brown LM, Elliott CG. When to offer genetic testing for pulmonary arterial hypertension. Can J Cardiol. 2015;31(4):544-7.
233. Jones DL, Clayton EW. The role of distress in uptake and response to predisposition genetic testing: the BMPR2 experience. Genet Test Mol Biomarkers. 2012;16(3):203-9.
234. Jones DL, Sandberg JC, Rosenthal MJ, Saunders RC, Hannig VL, Clayton EW. What patients and their relatives think about testing for *BMPR2*. J Genet Couns. 2008;17(5):452-8.
235. Girerd B, Montani D, Jais X, Eyries M, Yaici A, Sztrymf B, et al. Genetic counselling in a national referral centre for pulmonary hypertension. Eur Respir J. 2016;47(2):541-52.
236. Jacher JE, Martin LJ, Chung WK, Loyd JE, Nichols WC. Pulmonary arterial hypertension: Specialists' knowledge, practices, and attitudes of genetic counseling and genetic testing in the USA. Pulm Circ. 2017;7(2):372-83.
237. Morse JH, Barst RJ. Detection of familial primary pulmonary hypertension by genetic testing. N Engl J Med. 1997;337(3):202-3.

238. Song J, Eichstaedt CA, Viales RR, Benjamin N, Harutyunova S, Fischer C, et al. Identification of genetic defects in pulmonary arterial hypertension by a new gene panel diagnostic tool. Clin Sci (Lond). 2016;130(22):2043-52.
239. Higenbottam T, Butt AY, McMahon A, Westerbeck R, Sharples L. Long-term intravenous prostaglandin (epoprostenol or iloprost) for treatment of severe pulmonary hypertension. Heart. 1998;80(2):151-5.
240. Rubin LJ, Groves BM, Reeves JT, Frosolono M, Handel F, Cato AE. Prostacyclin-induced acute pulmonary vasodilation in primary pulmonary hypertension. Circulation. 1982;66(2):334-8.
241. Clapp LH, Gurung R. The mechanistic basis of prostacyclin and its stable analogues in pulmonary arterial hypertension: Role of membrane versus nuclear receptors. Prostaglandins Other Lipid Mediat. 2015;120:56-71.
242. Archer SL, Huang JM, Hampl V, Nelson DP, Shultz PJ, Weir EK. Nitric oxide and cGMP cause vasorelaxation by activation of a charybdotoxin-sensitive K channel by cGMP-dependent protein kinase. Proc Natl Acad Sci U S A. 1994;91(16):7583-7.
243. Galie N, Ghofrani HA, Torbicki A, Barst RJ, Rubin LJ, Badesch D, et al. Sildenafil citrate therapy for pulmonary arterial hypertension. N Engl J Med. 2005;353(20):2148-57.
244. Sastry BK, Narasimhan C, Reddy NK, Raju BS. Clinical efficacy of sildenafil in primary pulmonary hypertension: a randomized, placebo-controlled, double-blind, crossover study. J Am Coll Cardiol. 2004;43(7):1149-53.
245. Channick RN, Simonneau G, Sitbon O, Robbins IM, Frost A, Tapson VF, et al. Effects of the dual endothelin-receptor antagonist bosentan in patients with pulmonary hypertension: a randomised placebo-controlled study. Lancet. 2001;358(9288):1119-23.
246. Rubin LJ, Badesch DB, Barst RJ, Galie N, Black CM, Keogh A, et al. Bosentan therapy for pulmonary arterial hypertension. N Engl J Med. 2002;346(12):896-903.
247. Galie N, Rubin L, Hoeper M, Jansa P, Al-Hiti H, Meyer G, et al. Treatment of patients with mildly symptomatic pulmonary arterial hypertension with bosentan (EARLY study): a double-blind, randomised controlled trial. Lancet. 2008;371(9630):2093-100.
248. Galie N, Olschewski H, Oudiz RJ, Torres F, Frost A, Ghofrani HA, et al. Ambrisentan for the treatment of pulmonary arterial hypertension: results of the ambrisentan in pulmonary arterial hypertension, randomized, double-blind, placebo-controlled, multicenter, efficacy (ARIES) study 1 and 2. Circulation. 2008;117(23):3010-9.

249. Sitbon O, Channick R, Chin KM, Frey A, Gaine S, Galie N, et al. Selexipag for the Treatment of Pulmonary Arterial Hypertension. N Engl J Med. 2015;373(26):2522-33.
250. Pulido T, Adzerikho I, Channick RN, Delcroix M, Galie N, Ghofrani HA, et al. Macitentan and morbidity and mortality in pulmonary arterial hypertension. N Engl J Med. 2013;369(9):809-18.
251. Hoeper MM, Markevych I, Spiekerkoetter E, Welte T, Niedermeyer J. Goal-oriented treatment and combination therapy for pulmonary arterial hypertension. Eur Respir J. 2005;26(5):858-63.
252. McLaughlin VV, Gaine SP, Howard LS, Leuchte HH, Mathier MA, Mehta S, et al. Treatment goals of pulmonary hypertension. J Am Coll Cardiol. 2013;62(25 Suppl):D73-81.
253. Galie N, Barbera JA, Frost AE, Ghofrani HA, Hoeper MM, McLaughlin VV, et al. Initial use of ambrisentan plus tadalafil in pulmonary arterial hypertension. N Engl J Med. 2015;373(9):834-44.
254. Hoeper MM, McLaughlin VV, Barbera JA, Frost AE, Ghofrani HA, Peacock AJ, et al. Initial combination therapy with ambrisentan and tadalafil and mortality in patients with pulmonary arterial hypertension: a secondary analysis of the results from the randomised, controlled AMBITION study. Lancet Respir Med. 2016;4(11):894-901.
255. Sitbon O, Jais X, Savale L, Cottin V, Bergot E, Macari EA, et al. Upfront triple combination therapy in pulmonary arterial hypertension: a pilot study. Eur Respir J. 2014;43(6):1691-7.
256. Sandoval J, Gaspar J, Pulido T, Bautista E, Martinez-Guerra ML, Zeballos M, et al. Graded balloon dilation atrial septostomy in severe primary pulmonary hypertension. A therapeutic alternative for patients nonresponsive to vasodilator treatment. J Am Coll Cardiol. 1998;32(2):297-304.
257. Rothman A, Sklansky MS, Lucas VW, Kashani IA, Shaughnessy RD, Channick RN, et al. Atrial septostomy as a bridge to lung transplantation in patients with severe pulmonary hypertension. Am J Cardiol. 1999;84(6):682-6.
258. Chiu JS, Zuckerman WA, Turner ME, Richmond ME, Kerstein D, Krishnan U, et al. Balloon atrial septostomy in pulmonary arterial hypertension: effect on survival and associated outcomes. J Heart Lung Transplant. 2015;34(3):376-80.

259. Chen H, Shiboski SC, Golden JA, Gould MK, Hays SR, Hoopes CW, et al. Impact of the lung allocation score on lung transplantation for pulmonary arterial hypertension. Am J Respir Crit Care Med. 2009;180(5):468-74.
260. Schaffer JM, Singh SK, Joyce DL, Reitz BA, Robbins RC, Zamanian RT, et al. Transplantation for idiopathic pulmonary arterial hypertension: improvement in the lung allocation score era. Circulation. 2013;127(25):2503-13.
261. Gomberg-Maitland M, Glassner-Kolmin C, Watson S, Frantz R, Park M, Frost A, et al. Survival in pulmonary arterial hypertension patients awaiting lung transplantation. J Heart Lung Transplant. 2013;32(12):1179-86.
262. Keogh AM, Mayer E, Benza RL, Corris P, Dartevelle PG, Frost AE, et al. Interventional and surgical modalities of treatment in pulmonary hypertension. J Am Coll Cardiol. 2009;54(1 Suppl):S67-77.
263. Sarashina T, Nakamura K, Akagi S, Oto T, Oe H, Ejiri K, et al. Reverse right ventricular remodeling after lung transplantation in patients with pulmonary arterial hypertension under combination therapy of targeted medical drugs. Circ J. 2017;81(3):383-90.
264. Gorter TM, Verschuuren EAM, van Veldhuisen DJ, Hoendermis ES, Erasmus ME, Bogaard HJ, et al. Right ventricular recovery after bilateral lung transplantation for pulmonary arterial hypertensiondagger. Interact Cardiovasc Thorac Surg. 2017;24(6):890-7.
265. Yusen RD, Christie JD, Edwards LB, Kucheryavaya AY, Benden C, Dipchand AI, et al. The Registry of the International Society for Heart and Lung Transplantation: Thirtieth Adult Lung and Heart-Lung Transplant Report--2013. J Heart Lung Transplant. 2013;32(10):965-78.
266. Humbert M, Sitbon O, Yaici A, Montani D, O'Callaghan DS, Jais X, et al. Survival in incident and prevalent cohorts of patients with pulmonary arterial hypertension. Eur Respir J. 2010;36(3):549-55.
267. Simonneau G, Channick RN, Delcroix M, Galie N, Ghofrani HA, Jansa P, et al. Incident and prevalent cohorts with pulmonary arterial hypertension: insight from SERAPHIN. Eur Respir J. 2015;46(6):1711-20.
268. Brittain EL, Pugh ME, Wheeler LA, Robbins IM, Loyd JE, Newman JH, et al. Shorter survival in familial versus idiopathic pulmonary arterial hypertension is associated with hemodynamic markers of impaired right ventricular function. Pulm Circ. 2013;3(3):589-98.

269. Thenappan T, Shah SJ, Rich S, Tian L, Archer SL, Gomberg-Maitland M. Survival in pulmonary arterial hypertension: a reappraisal of the NIH risk stratification equation. Eur Respir J. 2010;35(5):1079-87.
270. Lee WT, Ling Y, Sheares KK, Pepke-Zaba J, Peacock AJ, Johnson MK. Predicting survival in pulmonary arterial hypertension in the UK. Eur Respir J. 2012;40(3):604-11.
271. Hoeper MM, Pletz MW, Golpon H, Welte T. Prognostic value of blood gas analyses in patients with idiopathic pulmonary arterial hypertension. Eur Respir J. 2007;29(5):944-50.
272. Zeng WJ, Sun YJ, Gu Q, Xiong CM, Li JJ, He JG. The impact of pulmonary arterial hypertension-targeted therapy on survival in Chinese patients with idiopathic pulmonary arterial hypertension. Pulm Circ. 2012;2(3):373-8.
273. Shimony A, Fox BD, Afilalo J, Rudski LG, Hirsch A, Langleben D. Pulmonary arterial hypertension in the elderly-clinical characteristics and long-term survival. Lung. 2012;190(6):645-9.
274. Shimony A, Eisenberg MJ, Rudski LG, Schlesinger R, Afilalo J, Joyal D, et al. Prevalence and impact of coronary artery disease in patients with pulmonary arterial hypertension. Am J Cardiol. 2011;108(3):460-4.
275. Shapiro S, Traiger GL, Turner M, McGoon MD, Wason P, Barst RJ. Sex differences in the diagnosis, treatment, and outcome of patients with pulmonary arterial hypertension enrolled in the registry to evaluate early and long-term pulmonary arterial hypertension disease management. Chest. 2012;141(2):363-73.
276. Escribano-Subias P, Blanco I, Lopez-Meseguer M, Lopez-Guarch CJ, Roman A, Morales P, et al. Survival in pulmonary hypertension in Spain: insights from the Spanish registry. Eur Respir J. 2012;40(3):596-603.
277. Ogawa A, Ejiri K, Matsubara H. Long-term patient survival with idiopathic/heritable pulmonary arterial hypertension treated at a single center in Japan. Life Sci. 2014;118(2):414-9.
278. Jansa P, Jarkovsky J, Al-Hiti H, Popelova J, Ambroz D, Zatocil T, et al. Epidemiology and long-term survival of pulmonary arterial hypertension in the Czech Republic: a retrospective analysis of a nationwide registry. BMC Pulm Med. 2014;14:45.
279. Benza RL, Miller DP, Barst RJ, Badesch DB, Frost AE, McGoon MD. An evaluation of long-term survival from time of diagnosis in pulmonary arterial hypertension from the REVEAL Registry. Chest. 2012;142(2):448-56.

280. Zhang R, Dai LZ, Xie WP, Yu ZX, Wu BX, Pan L, et al. Survival of Chinese patients with pulmonary arterial hypertension in the modern treatment era. Chest. 2011;140(2):301-9.
281. Montani D, Lau EM, Dorfmuller P, Girerd B, Jais X, Savale L, et al. Pulmonary veno-occlusive disease. Eur Respir J. 2016;47(5):1518-34.
282. Voordes CG, Kuipers JR, Elema JD. Familial pulmonary veno-occlusive disease: a case report. Thorax. 1977;32(6):763-6.
283. Langleben D, Heneghan JM, Batten AP, Wang NS, Fitch N, Schlesinger RD, et al. Familial pulmonary capillary hemangiomatosis resulting in primary pulmonary hypertension. Ann Intern Med. 1988;109(2):106-9.
284. Runo JR, Vnencak-Jones CL, Prince M, Loyd JE, Wheeler L, Robbins IM, et al. Pulmonary veno-occlusive disease caused by an inherited mutation in bone morphogenetic protein receptor II. Am J Respir Crit Care Med. 2003;167(6):889-94.
285. Machado RD, Aldred MA, James V, Harrison RE, Patel B, Schwalbe EC, et al. Mutations of the TGF-beta type II receptor *BMPR2* in pulmonary arterial hypertension. Hum Mutat. 2006;27(2):121-32.
286. Zhang P, McGrath BC, Reinert J, Olsen DS, Lei L, Gill S, et al. The GCN2 eIF2alpha kinase is required for adaptation to amino acid deprivation in mice. Mol Cell Biol. 2002;22(19):6681-8.
287. Dong J, Qiu H, Garcia-Barrio M, Anderson J, Hinnebusch AG. Uncharged tRNA activates GCN2 by displacing the protein kinase moiety from a bipartite tRNA-binding domain. Mol Cell. 2000;6(2):269-79.
288. Dever TE, Feng L, Wek RC, Cigan AM, Donahue TF, Hinnebusch AG. Phosphorylation of initiation factor 2 alpha by protein kinase GCN2 mediates gene-specific translational control of GCN4 in yeast. Cell. 1992;68(3):585-96.
289. Holcik M, Sonenberg N. Translational control in stress and apoptosis. Nat Rev Mol Cell Biol. 2005;6(4):318-27.
290. Krishnamoorthy T, Pavitt GD, Zhang F, Dever TE, Hinnebusch AG. Tight binding of the phosphorylated alpha subunit of initiation factor 2 (eIF2alpha) to the regulatory subunits of guanine nucleotide exchange factor eIF2B is required for inhibition of translation initiation. Mol Cell Biol. 2001;21(15):5018-30.
291. Vattem KM, Wek RC. Reinitiation involving upstream ORFs regulates ATF4 mRNA translation in mammalian cells. Proc Natl Acad Sci U S A. 2004;101(31):11269-74.

292. Perros F, Gunther S, Ranchoux B, Godinas L, Antigny F, Chaumais MC, et al. Mitomycin-induced pulmonary veno-occlusive disease: evidence from human disease and animal models. Circulation. 2015;132(9):834-47.
293. Ranchoux B, Gunther S, Quarck R, Chaumais MC, Dorfmuller P, Antigny F, et al. Chemotherapy-induced pulmonary hypertension: role of alkylating agents. Am J Pathol. 2015;185(2):356-71.
294. Montani D, Lau EM, Descatha A, Jais X, Savale L, Andujar P, et al. Occupational exposure to organic solvents: a risk factor for pulmonary veno-occlusive disease. Eur Respir J. 2015;46(6):1721-31.
295. Lantuejoul S, Sheppard MN, Corrin B, Burke MM, Nicholson AG. Pulmonary veno-occlusive disease and pulmonary capillary hemangiomatosis: a clinicopathologic study of 35 cases. Am J Surg Pathol. 2006;30(7):850-7.
296. Dorfmuller P, Humbert M, Perros F, Sanchez O, Simonneau G, Muller KM, et al. Fibrous remodeling of the pulmonary venous system in pulmonary arterial hypertension associated with connective tissue diseases. Hum Pathol. 2007;38(6):893-902.
297. Overbeek MJ, Vonk MC, Boonstra A, Voskuyl AE, Vonk-Noordegraaf A, Smit EF, et al. Pulmonary arterial hypertension in limited cutaneous systemic sclerosis: a distinctive vasculopathy. Eur Respir J. 2009;34(2):371-9.
298. Rabiller A, Jais X, Hamid A, Resten A, Parent F, Haque R, et al. Occult alveolar haemorrhage in pulmonary veno-occlusive disease. Eur Respir J. 2006;27(1):108-13.
299. Pietra GG, Edwards WD, Kay JM, Rich S, Kernis J, Schloo B, et al. Histopathology of primary pulmonary hypertension. A qualitative and quantitative study of pulmonary blood vessels from 58 patients in the National Heart, Lung, and Blood Institute, Primary Pulmonary Hypertension Registry. Circulation. 1989;80(5):1198-206.
300. Montani D, Achouh L, Dorfmuller P, Le Pavec J, Sztrymf B, Tcherakian C, et al. Pulmonary veno-occlusive disease: clinical, functional, radiologic, and hemodynamic characteristics and outcome of 24 cases confirmed by histology. Medicine (Baltimore). 2008;87(4):220-33.
301. Wagenvoort CA. Pulmonary veno-occlusive disease. Entity or syndrome? Chest. 1976;69(1):82-6.
302. Holcomb BW, Jr., Loyd JE, Ely EW, Johnson J, Robbins IM. Pulmonary veno-occlusive disease: a case series and new observations. Chest. 2000;118(6):1671-9.

303. Elliott CG, Colby TV, Hill T, Crapo RO. Pulmonary veno-occlusive disease associated with severe reduction of single-breath carbon monoxide diffusing capacity. Respiration. 1988;53(4):262-6.
304. Godinas L, Amar D, Montani D, Lau EM, Jais X, Savale L, et al. Lung capillary blood volume and membrane diffusion in precapillary pulmonary hypertension. J Heart Lung Transplant. 2016;35(5):647-56.
305. Palevsky HI, Pietra GG, Fishman AP. Pulmonary veno-occlusive disease and its response to vasodilator agents. Am Rev Respir Dis. 1990;142(2):426-9.
306. Montani D, Girerd B, Jais X, Levy M, Amar D, Savale L, et al. Clinical phenotypes and outcomes of heritable and sporadic pulmonary veno-occlusive disease: a population-based study. Lancet Respir Med. 2017;5(2):125-34.
307. Creagh-Brown BC, Nicholson AG, Showkathali R, Gibbs JS, Howard LS. Pulmonary veno-occlusive disease presenting with recurrent pulmonary oedema and the use of nitric oxide to predict response to sildenafil. Thorax. 2008;63(10):933-4.
308. Resten A, Maitre S, Humbert M, Rabiller A, Sitbon O, Capron F, et al. Pulmonary hypertension: CT of the chest in pulmonary venoocclusive disease. AJR Am J Roentgenol. 2004;183(1):65-70.
309. Kuroda T, Hirota H, Masaki M, Sugiyama S, Oshima Y, Terai K, et al. Sildenafil as adjunct therapy to high-dose epoprostenol in a patient with pulmonary veno-occlusive disease. Heart Lung Circ. 2006;15(2):139-42.
310. Ogawa A, Miyaji K, Yamadori I, Shinno Y, Miura A, Kusano KF, et al. Safety and efficacy of epoprostenol therapy in pulmonary veno-occlusive disease and pulmonary capillary hemangiomatosis. Circ J. 2012;76(7):1729-36.
311. Palmer SM, Robinson LJ, Wang A, Gossage JR, Bashore T, Tapson VF. Massive pulmonary edema and death after prostacyclin infusion in a patient with pulmonary veno-occlusive disease. Chest. 1998;113(1):237-40.
312. Humbert M, Maitre S, Capron F, Rain B, Musset D, Simonneau G. Pulmonary edema complicating continuous intravenous prostacyclin in pulmonary capillary hemangiomatosis. Am J Respir Crit Care Med. 1998;157(5 Pt 1):1681-5.
313. Montani D, Jais X, Price LC, Achouh L, Degano B, Mercier O, et al. Cautious epoprostenol therapy is a safe bridge to lung transplantation in pulmonary veno-occlusive disease. Eur Respir J. 2009;34(6):1348-56.

314. Wille KM, Sharma NS, Kulkarni T, Lammi MR, Barney JB, Bellot SC, et al. Characteristics of patients with pulmonary venoocclusive disease awaiting transplantation. Ann Am Thorac Soc. 2014;11(9):1411-8.
315. The Cost of Sequencing a Human Genome 2017 [Available from: https://www.genome.gov/27565109/The-Cost-of-Sequencing-a-Human-Genome.
316. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, et al. Accurate whole human genome sequencing using reversible terminator chemistry. Nature. 2008;456(7218):53-9.
317. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet. 2016;17(6):333-51.
318. Cohen JC, Kiss RS, Pertsemlidis A, Marcel YL, McPherson R, Hobbs HH. Multiple rare alleles contribute to low plasma levels of HDL cholesterol. Science. 2004;305(5685):869-72.
319. Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, et al. Targeted capture and massively parallel sequencing of 12 human exomes. Nature. 2009;461(7261):272-6.
320. Meienberg J, Bruggmann R, Oexle K, Matyas G. Clinical sequencing: is WGS the better WES? Hum Genet. 2016;135(3):359-62.
321. Meienberg J, Zerjavic K, Keller I, Okoniewski M, Patrignani A, Ludin K, et al. New insights into the performance of human whole-exome capture platforms. Nucleic Acids Res. 2015;43(11):e76.
322. Belkadi A, Bolze A, Itan Y, Cobat A, Vincent QB, Antipenko A, et al. Whole-genome sequencing is more powerful than whole-exome sequencing for detecting exome variants. Proc Natl Acad Sci U S A. 2015;112(17):5473-8.
323. Next Generation Genomics: World Map of High-throughput Sequencers 2017 [Available from: http://omicsmaps.com/.
324. Medvedev P, Stanciu M, Brudno M. Computational methods for discovering structural variation with next-generation sequencing. Nat Methods. 2009;6(11 Suppl):S13-20.
325. Moorthie S, Hall A, Wright CF. Informatics and clinical genome sequencing: opening the black box. Genet Med. 2013;15(3):165-71.

326. Guo Y, Dai Y, Yu H, Zhao S, Samuels DC, Shyr Y. Improvements and impacts of GRCh38 human reference on high throughput sequencing data analysis. Genomics. 2017;109(2):83-90.
327. Trapnell C, Salzberg SL. How to map billions of short reads onto genomes. Nat Biotechnol. 2009;27(5):455-7.
328. Altmann A, Weber P, Bader D, Preuss M, Binder EB, Muller-Myhsok B. A beginners guide to SNP calling from high-throughput DNA-sequencing data. Hum Genet. 2012;131(10):1541-54.
329. Nielsen R, Paul JS, Albrechtsen A, Song YS. Genotype and SNP calling from next-generation sequencing data. Nat Rev Genet. 2011;12(6):443-51.
330. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016;536(7616):285-91.
331. Walter K, Min JL, Huang J, Crooks L, Memari Y, McCarthy S, et al. The UK10K project identifies rare variants in health and disease. Nature. 2015;526(7571):82-90.
332. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet. 2014;46(3):310-5.
333. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nat Methods. 2010;7(4):248-9.
334. Ng PC, Henikoff S. Predicting deleterious amino acid substitutions. Genome Res. 2001;11(5):863-74.
335. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The Ensembl Variant Effect Predictor. Genome Biol. 2016;17(1):122.
336. Zarrei M, MacDonald JR, Merico D, Scherer SW. A copy number variation map of the human genome. Nat Rev Genet. 2015;16(3):172-83.
337. Staples J, Qiao D, Cho M, Silverman E, Nickerson D, Below J. PRIMUS: Rapid Reconstruction of Pedigrees from Genome-wide Estimates of Identity by Descent. Am J Hum Genet. 2014;95(5):553-64.
338. Hothorn T, Hornik K, Wiel MAvd, Zeileis A. A Lego System for Conditional Inference. The American Statistician. 2012;60(3):257-63.

339. Kloke JD, McKean JW. Rfit: Rank-based estimation for linear models. The R Journal. 2012;4(2):57-64.
340. Therneau T, Grambsch P. Modeling survival data: extending the Cox model. 1 ed. New York: Springer-Verlag 2000. 350 p.
341. Grambsch P, Therneau H. Proportional hazards tests and diagnostics based on weighted residuals. Biometrika. 1994;81(3):515-26.
342. Collett D. Modelling survival data in medical research. 3rd ed. London: Chapman & Hall/CRC; 2014.
343. Rubenfire M, Huffman MD, Krishnan S, Seibold JR, Schiopu E, McLaughlin VV. Survival in systemic sclerosis with pulmonary arterial hypertension has not improved in the modern era. Chest. 2013;144(4):1282-90.
344. Simonneau G, Fartoukh M, Sitbon O, Humbert M, Jagot JL, Herve P. Primary pulmonary hypertension associated with the use of fenfluramine derivatives. Chest. 1998;114(3 Suppl):195s-9s.
345. EudraLex - Volume 9 - Pharmacovigilance guidelines - Public Health - European Commission: European Commission; 2017 [Available from: https://ec.europa.eu/health/documents/eudralex/vol-9_en.
346. Taichman DB, McGoon MD, Harhay MO, Archer-Chicko C, Sager JS, Murugappan M, et al. Wide variation in clinicians' assessment of New York Heart Association/World Health Organization functional class in patients with pulmonary arterial hypertension. Mayo Clin Proc. 2009;84(7):586-92.
347. Pugh ME, Hemnes AR, Trammell A, Newman JH, Robbins IM. Variability in hemodynamic evaluation of pulmonary hypertension at large referral centers. Pulm Circ. 2014;4(4):679-84.
348. McLaughlin VV, Suissa S. Prognosis of pulmonary arterial hypertension: the power of clinical registries of rare diseases. Circulation. 2010;122(2):106-8.
349. Benza RL, Gomberg-Maitland M, Miller DP, Frost A, Frantz RP, Foreman AJ, et al. The REVEAL Registry risk score calculator in patients newly diagnosed with pulmonary arterial hypertension. Chest. 2012;141(2):354-62.
350. Sitbon O, Benza RL, Badesch DB, Barst RJ, Elliott CG, Gressin V, et al. Validation of two predictive models for survival in pulmonary arterial hypertension. Eur Respir J. 2015;46(1):152-64.

351. Trip P, Nossent EJ, de Man FS, van den Berk IA, Boonstra A, Groepenhoff H, et al. Severely reduced diffusion capacity in idiopathic pulmonary arterial hypertension: patient characteristics and treatment responses. Eur Respir J. 2013;42(6):1575-85.
352. Ragosta M. Textbook of Clinical Hemodynamics. 2nd ed. Philadelphia, USA: Elsevier; 2018. 328 p.
353. Drake KM, Zygmunt D, Mavrakis L, Harbor P, Wang L, Comhair SA, et al. Altered MicroRNA processing in heritable pulmonary arterial hypertension: an important role for Smad-8. Am J Respir Crit Care Med. 2011;184(12):1400-8.
354. Cummings J, Masten J. Customized dual data entry for computerized data analysis. Qual Assur. 1994;3(3):300-3.
355. Arnold AM, Kronmal RA. Multiple imputation of baseline data in the cardiovascular health study. Am J Epidemiol. 2003;157(1):74-84.
356. Clinical Audits and Registries Management Service ND. National Audit of Pulmonary Hypertension - 7th Annual Report: The Health and Social Care Information Centre; 2017 [Available from: http://www.content.digital.nhs.uk/catalogue/PUB23648.
357. Chandra S, Shah SJ, Thenappan T, Archer SL, Rich S, Gomberg-Maitland M. Carbon monoxide diffusing capacity and mortality in pulmonary arterial hypertension. J Heart Lung Transplant. 2010;29(2):181-7.
358. Allanore Y, Borderie D, Avouac J, Zerkak D, Meune C, Hachulla E, et al. High N-terminal pro-brain natriuretic peptide levels and low diffusing capacity for carbon monoxide as independent predictors of the occurrence of precapillary pulmonary arterial hypertension in patients with systemic sclerosis. Arthritis Rheum. 2008;58(1):284-91.
359. van der Bruggen CE, Spruijt OA, Nossent EJ, Trip P, Marcus JT, de Man FS, et al. Treatment response in patients with idiopathic pulmonary arterial hypertension and a severely reduced diffusion capacity. Pulm Circ. 2017;7(1):137-44.
360. Trip P, Girerd B, Bogaard HJ, de Man FS, Boonstra A, Garcia G, et al. Diffusion capacity and *BMPR2* mutations in pulmonary arterial hypertension. Eur Respir J. 43. England2014. p. 1195-8.
361. GTEx Portal. *BMPR2* gene expression. 2018 [Available from: https://www.gtexportal.org/home/gene/BMPR2.

362. Lambert M, Boet A, Rucker-Martin C, Mendes-Ferreira P, Capuano V, Hatem S, et al. Loss of *KCNK3* is a hallmark of RV hypertrophy/dysfunction associated with pulmonary hypertension. Cardiovasc Res. 2018.
363. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov. 2012;2(5):401-4.
364. ARUP Scientific Resource for Research and Education: HHT Disease Databases | University of Utah 2018 [Available from: http://www.arup.utah.edu/database/HHT/.
365. Mayeur C, Leyton PA, Kolodziej SA, Yu B, Bloch KD. BMP type II receptors have redundant roles in the regulation of hepatic hepcidin gene expression and iron metabolism. Blood. 2014;124(13):2116-23.
366. Girerd B, Coulet F, Jais X, Eyries M, Van Der Bruggen C, De Man F, et al. Characteristics of pulmonary arterial hypertension in affected carriers of a mutation located in the cytoplasmic tail of bone morphogenetic protein receptor type 2. Chest. 2015;147(5):1385-94.
367. John A, Kizhakkedath P, Al-Gazali L, Ali BR. Defective cellular trafficking of the bone morphogenetic protein receptor type II by mutations underlying familial pulmonary arterial hypertension. Gene. 2015;561(1):148-56.
368. Mandel J, Mark EJ, Hales CA. Pulmonary veno-occlusive disease. Am J Respir Crit Care Med. 2000;162(5):1964-73.
369. Pietra GG, Edwards WD, Kay JM, Rich S, Kernis J, Schloo B, et al. Histopathology of primary pulmonary hypertension. A qualitative and quantitative study of pulmonary blood vessels from 58 patients in the National Heart, Lung, and Blood Institute, Primary Pulmonary Hypertension Registry. Circulation. 1989;80(5):1198-206.
370. Harding HP, Zhang Y, Zeng H, Novoa I, Lu PD, Calfon M, et al. An integrated stress response regulates amino acid metabolism and resistance to oxidative stress. Mol Cell. 2003;11(3):619-33.
371. Eichstaedt CA, Song J, Benjamin N, Harutyunova S, Fischer C, Grunig E, et al. *EIF2AK4* mutation as "second hit" in hereditary pulmonary arterial hypertension. Respir Res. 2016;17(1):141.
372. Villaschi S, Pietra GG. Alveolo-capillary membrane in primary pulmonary hypertension. Appl Pathol. 1986;4(3):132-7.

373. Graf S, Haimel M, Bleda M, Hadinnapola C, Southgate L, Li W, et al. Identification of rare sequence variation underlying heritable pulmonary arterial hypertension. Nat Commun. 2018;9(1):1416.
374. Guazzi M. Pulmonary hypertension in heart failure preserved ejection fraction: prevalence, pathophysiology, and clinical perspectives. Circ Heart Fail. 2014;7(2):367-77.
375. Chemla D, Weatherald J, Lau EMT, Savale L, Boucly A, Attal P, et al. Clinical and hemodynamic correlates of pulmonary arterial stiffness in incident, untreated patients with idiopathic pulmonary arterial hypertension. Chest. 2018.
376. Crawley SF, Johnson MK, Dargie HJ, Peacock AJ. LA volume by CMR distinguishes idiopathic from pulmonary hypertension due to HFpEF. JACC Cardiovasc Imaging. 2013;6(10):1120-1.
377. Fayyaz AU, Edwards WD, Maleszewski JJ, Konik EA, DuBrock HM, Borlaug BA, et al. Global pulmonary vascular remodeling in pulmonary hypertension associated with heart failure and preserved or reduced ejection fraction. Circulation. 2018;137(17):1796-810.
378. Burgel PR, Paillasseur JL, Roche N. Identification of clinical phenotypes using cluster analyses in COPD patients with multiple comorbidities. Biomed Res Int. 2014;2014:420134.
379. Newby C, Heaney LG, Menzies-Gow A, Niven RM, Mansur A, Bucknall C, et al. Statistical cluster analysis of the British Thoracic Society Severe refractory Asthma Registry: clinical outcomes and phenotype stability. PLoS One. 2014;9(7):e102987.

Appendices

Appendix 1

Recruiting centres for the NIHR BRIDGE PAH Study		
Country	**Centre**	**Principle investigator**
United Kingdom	Freeman Hospital, Newcastle	Paul A Corris
	Golden Jubilee National Hospital, Glasgow	Andrew Peacock
	Great Ormond Street Hospital, London	Shahin Moledina
	Hammersmith Hospital and Imperial College, London	Martin R Wilkins
	Papworth Hospital, Cambridge	Joanna Pepke-Zaba
	Royal Brompton Hospital, London	Stephen J Wort
	Royal Free Hospital, London	Gerry Coghlan
	Royal Hallamshire Hospital, Sheffield	David G Kiely
	Royal United Hospitals Bath NHS Foundation Trust, Bath	Jay Suntharalingam
France	University of South Paris, Paris	Marc Humbert
Germany	University Hospital Giessen, Giessen	Henning Gall
Italy	San Matteo Hospital, Pavia	Stefano Ghio
Netherlands	VU University Medical Center, Amsterdam	Anton Vonk Noordegraaf

Appendix 2

Rare disease cohorts recruiting to the NIHR BioResource - Rare Diseases Study
Genomics England
Specialist Pathology: Evaluating Exomes in Diagnostics
Primary Immune Disorders
Bleeding and Platelet Disorders
Pulmonary Arterial Hypertension
Multiple Primary Malignant Tumours
Hypertrophic Cardiomyopathy
Cerebral Small Vessel Diseases
Steroid Resistant Nephrotic Syndrome
Intrahepatic Cholestasis of Pregnancy
Stem Cell & Myeloid Disorders
Primary Membranoproliferative Glomerulonephritis
Neuropathic Pain Disorder
Leber Hereditary Optic Neuropathy
Control
Ehlers-Danlos Syndromes

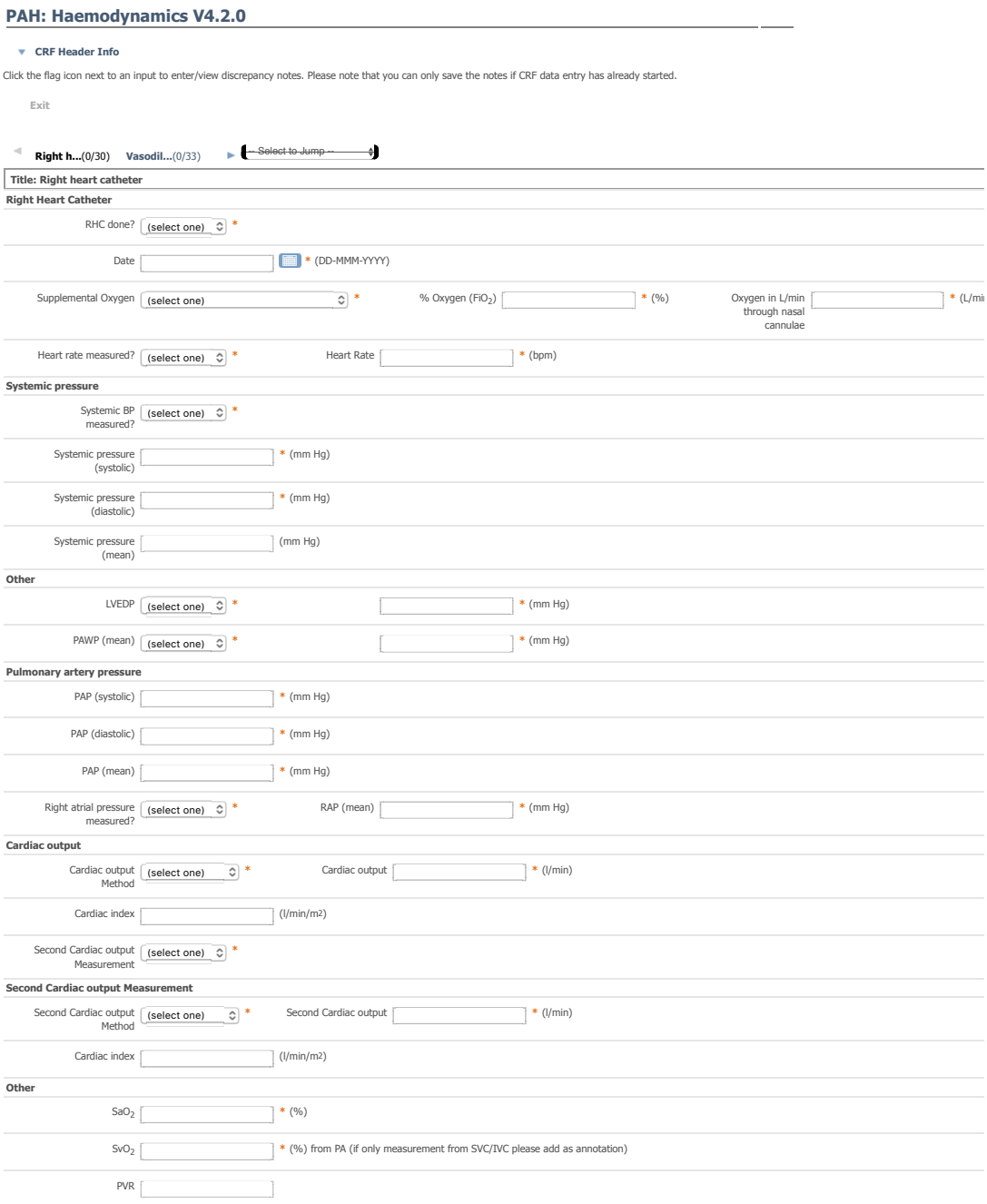
Appendix 3

Example of a template used to create OpenClinica eCRFs

ITEM NAME	DESCRIPTION LABEL	LEFT ITEM TEXT	UNITS	RIGHT ITEM TEXT	SECTION LABEL	GROUP LABEL	HEADER	SUBHEADER	PARENT ITEM	COLUMN NUMBER	PAGE NUMBER	QUESTION NUMBER	RESPONSE TYPE	RESPONSE LABEL	RESPONSE OPTIONS TEXT	RESPONSE VALUES OR CALCULATIONS	RESPONSE LAYOUT	DEFAULT VALUE	DATA TYPE	WIDTH DECIMAL	VALIDATION	VALIDATION ERROR MESSAGE	PHI	REQUIRED	ITEM DISPLAY STATUS	SIMPLE CONDITIONAL DISPLAY
HB_Study	Baseline study	RHC done?			Right heart catheter		Right Heart Catheter			1			single-select	D_N	(select one),done,not done	,done,not done		(select one)	ST				0	1	SHOW	
HB_Date	Date	Date	DD-MMM-YYYY		Right heart catheter					1			text	date					DATE				0	1	HIDE	HB_Study,done,ERROR
HB_SUPPLEMENTAL_OXYGEN	Supplemental Oxygen	Supplemental Oxygen			Right heart catheter					1			single-select	N_PC_Lmin	(select one),No,Unknown,% (through Venturi mask), L/min (for other forms of O2 delivery)	,no,UNK,%,Lmin			ST				0	1	HIDE	HB_Study,done,ERROR
HB_SUPPLEMENTAL_OXYGEN_LEVEL	% Oxygen (FiO2)	% Oxygen (FiO₂)	%		Right heart catheter					2			text	real					REAL				0	1	HIDE	HB_SUPPLEMENTAL_OXYGEN,%,ERROR
HB_SUPPLEMENTAL_OXYGEN_LEVEL_2	L/min Oxygen	Oxygen in L/min through nasal cannulae	L/min		Right heart catheter					2			text	real					REAL				0	1	HIDE	HB_SUPPLEMENTAL_OXYGEN,Lmin,ERROR
HB_HR_DONE	HR done	Heart rate measured?			Right heart catheter					1			single-select	D_N	(select one),done,not done	,done,not done		(select one)	ST				0	1	HIDE	HB_Study,done,ERROR
HB_HEART_RATE	Heart Rate	Heart Rate	bpm		Right heart catheter					2			text	real					REAL				0	1	HIDE	HB_HR_DONE,done,ERROR
HB_BP_DONE	Systemic BP done	Systemic BP measured?			Right heart catheter		Systemic pressure			1			single-select	D_N	(select one),done,not done	,done,not done		(select one)	ST				0	1	HIDE	HB_Study,done,ERROR
HB_SP_SYST	Systemic pressure: Systolic	Systemic pressure (systolic)	mm Hg		Right heart catheter								text	real					REAL				0	1	HIDE	HB_BP_DONE,done,ERROR
HB_SP_DIAS	Systemic pressure: Diastolic	Systemic pressure (diastolic)	mm Hg		Right heart catheter								text	real					REAL				0	1	HIDE	HB_BP_DONE,done,ERROR
HB_SP_MEAN	Systemic pressure: Mean	Systemic pressure (mean)	mm Hg		Right heart catheter								calculation	CALC_PRESSURE_HB	calculation	func: ((HB_SP_SYST+HB_SP_DIAS+HB_SP_DIAS)/3)			REAL				0	0	HIDE	
HB_LVEDP_DONE	LVEDP select	LVEDP			Right heart catheter		Other			1			single-select	ND_mmHg	(select one),not done,mm Hg	,not done,mm Hg		(select one)	ST				0	1	HIDE	HB_Study,done,ERROR
HB_LVEDP	LVEDP		mm Hg		Right heart catheter					2			text	real					REAL				0	1	HIDE	HB_LVEDP_DONE,mm Hg,ERROR
HB_PAWP_M_DONE	PAWP: mean select	PAWP (mean)			Right heart catheter					1			single-select	ND_mmHg				(select one)	ST				0	1	HIDE	HB_Study,done,ERROR
HB_PAWP_M	PAWP: mean		mm Hg		Right heart catheter					2			text	real					REAL				0	1	HIDE	HB_PAWP_M_DONE,mm Hg,ERROR
HB_PAP_S	PAP: Systolic	PAP (systolic)	mm Hg		Right heart catheter		Pulmonary artery pressure						text	real					REAL				0	1	HIDE	HB_Study,done,ERROR
HB_PAP_D	PAP: Diastolic	PAP (diastolic)	mm Hg		Right heart catheter								text	real					REAL				0	1	HIDE	HB_Study,done,ERROR
HB_PAP_M	PAP: mean	PAP (mean)	mm Hg		Right heart catheter								text	real					REAL				0	1	HIDE	HB_Study,done,ERROR
HB_RAP_DONE	RAP done	Right atrial pressure measured?			Right heart catheter					1			single-select	D_N	(select one),done,not done	,done,not done		(select one)	ST				0	1	HIDE	HB_Study,done,ERROR
HB_RAP_M	RAP: mean	RAP (mean)	mm Hg		Right heart catheter		Right atrial pressure			2			text	real					REAL				0	1	HIDE	HB_RAP_DONE,done,ERROR
HB_CARDIAC_OUTPUT_METHOD_1	Cardiac output Method	Cardiac output Method			Right heart catheter		Cardiac output			1			single-select	A_M_T_Ox	(select one),assumed Fick, measured Fick, thermodilution, other	,assumed,measured,thermodilution,other		(select one)	ST				0	1	HIDE	HB_Study,done,ERROR
HB_CARDIAC_OUTPUT_VALUE_1	Cardiac output	Cardiac output	l/min		Right heart catheter					2			text	real					REAL				0	1	HIDE	HB_Study,done,ERROR
HB_CARDIAC_INDEX_VALUE_1	Cardiac index	Cardiac index	l/min/m²		Right heart catheter					1			text	real					REAL				0	0	HIDE	HB_Study,done,ERROR
HB_CARDIAC_OUTPUT_2	Second Cardiac output Measurement	Second Cardiac output Measurement			Right heart catheter					1			single-select	Y_N-ss	(select one),Yes,No,Unkown	,yes,no,UNK		(select one)	ST				0	1	HIDE	HB_Study,done,ERROR
HB_CARDIAC_OUTPUT_METHOD_2	Second Cardiac output Method	Second Cardiac output Method			Right heart catheter		Second Cardiac output Measurement			1			single-select	A_M_T_Ox				(select one)	ST				0	1	HIDE	HB_CARDIAC_OUTPUT_2,yes,ERROR
HB_CARDIAC_OUTPUT_VALUE_2	Second Cardiac output	Second Cardiac output	l/min		Right heart catheter					2			text	real					REAL				0	1	HIDE	HB_CARDIAC_OUTPUT_2,yes,ERROR
HB_CARDIAC_INDEX_VALUE_2	Second Cardiac index	Cardiac index	l/min/m²		Right heart catheter					1			text	real					REAL				0	0	HIDE	HB_Study,done,ERROR
HB_SA_O2	SaO2	SaO₂	%		Right heart catheter		Other						text	real					REAL				0	1	HIDE	HB_Study,done,ERROR
HB_SV_O2	SvO2	SvO₂	%	from PA (if only measurement from SVC/IVC please add as annotation)	Right heart catheter								text	real					REAL				0	1	HIDE	HB_Study,done,ERROR
HB_PVR_Calc	PVR	PVR			Right heart catheter								calculation	PVR_Calc	calculation	func: (80*((HB_PAP_M - HB_PAWP_M)/HB_CARDIAC_OUTPUT_VALUE_1))			REAL				0	0	HIDE	

Appendix 4

Screenshot of the eCRF used to capture haemodynamic data in OpenClinica.



PAH: Haemodynamics V4.2.0

CRF Header Info

Click the flag icon next to an input to enter/view discrepancy notes. Please note that you can only save the notes if CRF data entry has already started.

Exit

Right h...(0/30) Vasodil...(0/33) -- Select to Jump --

Title: Right heart catheter

Right Heart Catheter

RHC done? (select one) *

Date * (DD-MMM-YYYY)

Supplemental Oxygen (select one) * % Oxygen (FiO_2) * (%) Oxygen in L/min through nasal cannulae * (L/mi

Heart rate measured? (select one) * Heart Rate * (bpm)

Systemic pressure

Systemic BP measured? (select one) *

Systemic pressure (systolic) * (mm Hg)

Systemic pressure (diastolic) * (mm Hg)

Systemic pressure (mean) (mm Hg)

Other

LVEDP (select one) * * (mm Hg)

PAWP (mean) (select one) * * (mm Hg)

Pulmonary artery pressure

PAP (systolic) * (mm Hg)

PAP (diastolic) * (mm Hg)

PAP (mean) * (mm Hg)

Right atrial pressure measured? (select one) * RAP (mean) * (mm Hg)

Cardiac output

Cardiac output Method (select one) * Cardiac output * (l/min)

Cardiac index (l/min/m2)

Second Cardiac output Measurement (select one) *

Second Cardiac output Measurement

Second Cardiac output Method (select one) * Second Cardiac output * (l/min)

Cardiac index (l/min/m2)

Other

SaO_2 * (%)

SvO_2 * (%) from PA (if only measurement from SVC/IVC please add as annotation)

PVR

Appendix 5

Discrepancy notes

When adding a Discrepancy note (or Flag) to the data you are entering, the Type of Note should be determined by the reason you are adding the note. **Please use "Reason for Change", "Annotation", "Query" or "Failed Validation Check" appropriately**.

1) Please select the correct "*Type*" of discrepancy note.
2) If additional information is available / data updated or amended in OC / an error is corrected etc. then select "**Reason for Change**".
3) Only select "**Annotation**" type discrepancy notes if you feel that information is available that should be recorded in OC but you cannot input the data appropriately. (See below for further information).

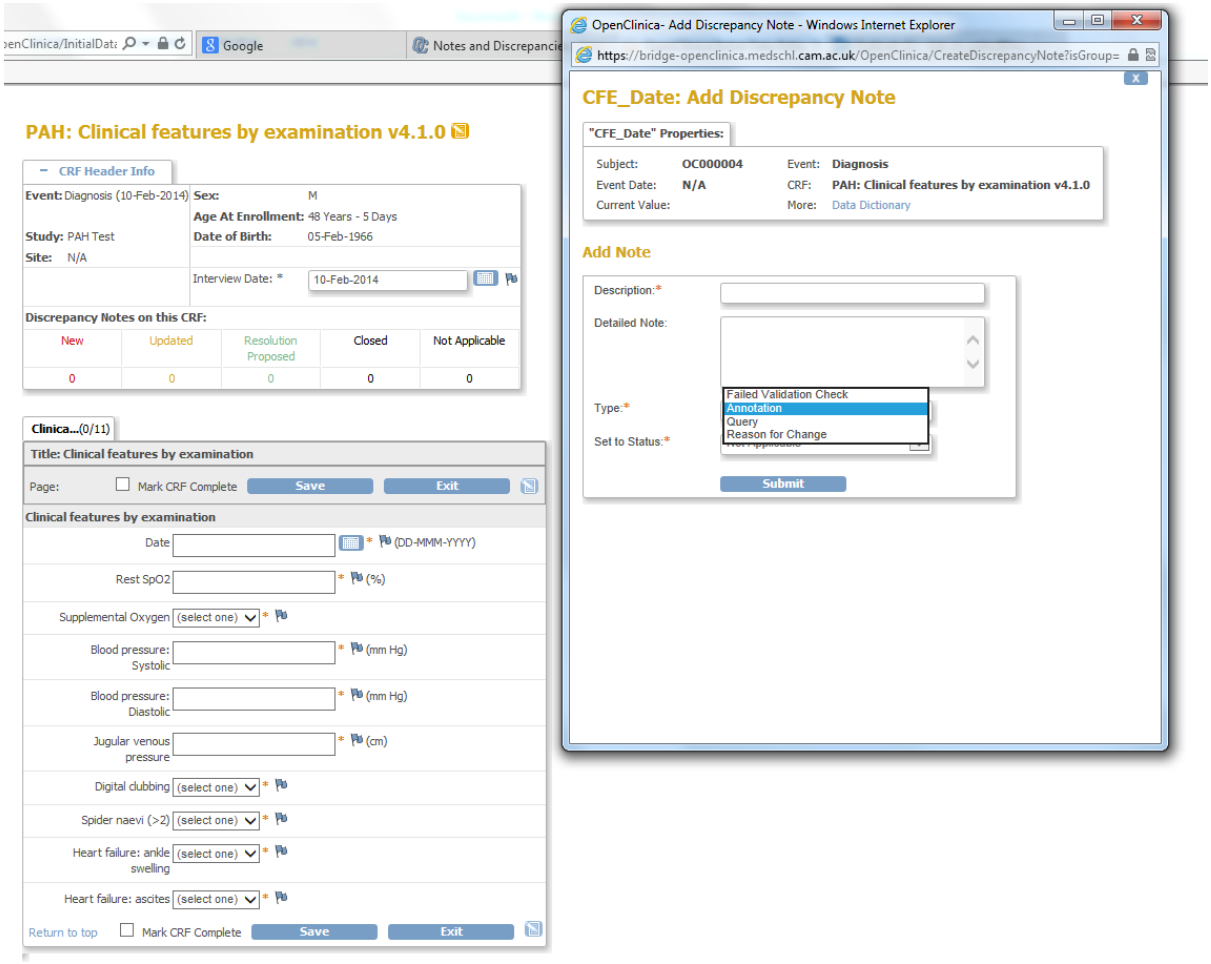
'Reason for Change' Discrepancy Notes

1) To speed up discrepancy note entry and analysis please use the following controlled vocabulary in the "Description" section of the discrepancy note:
   a) If new information is available please state "**UPDATE**" in the discrepancy note description. Further information can be provided in the details section if needed.
   b) If new information is added following a request during a monitoring visit please state "**MONITORING**" in the discrepancy note description. Further information can be provided in the details section if needed.
   c) If new information is added following a data/discrepancy note query please state "**RESPONSE**" in the discrepancy note description. Further information can be provided in the details section if needed.
2) If you think further clinical information / justification is important and needs to be added to the discrepancy note please add this into the "detailed notes" section.

'Annotation' Discrepancy Notes

Annotations are used to make comments or provide information about the data that cannot be adequately represented in the CRF. For example if data is definitely missing or is in a different format that cannot be converted accurately.

These types of notes are not intended to begin a conversation thread but are to be used to record data which otherwise cannot be entered into a field. Where data is missing from a non-compulsory field (* indicates compulsory), a note to explain that data is missing is not required.



Again use the controlled vocabulary in the "description" box and any additional information in the "detailed notes" section. Please use one of these 3 words in the "description" field:

i) If an annotation is flagging up a clinical finding (e.g. MDT outcome) please state "**CLINICAL**". Please expand in the "detailed notes" section.

ii) If an annotation is flagging up uncertainty in the data (e.g. a date) entered into OC please state "**UNCERTAINTY**". Please expand in the detailed notes section.

iii) If an annotation is there to justify an unexpected result please state "**JUSTIFY**". For example a lab result for which there is no reference range. Please expand in the "detailed notes" section. Note this is different to the "Failed Validation Check" type discrepancy note (see below).

"Failed Validation Check"

A Failed Validation Check Note Type is for data that does not comply with expected values. For example where a range has been set on a data field but in this particular case the patient's reading is recorded as out of range and you are sure this is correct.

These types of notes are not intended to begin a conversation thread but are to be used to record data which otherwise cannot be entered into a field. This note can be created manually, useful if you wish to add a detailed note. Otherwise if out of range data is entered and the page saved, OpenClinica first displays a warning message, then if the value is not changed to within range it creates a Failed Validation Check Discrepancy Note automatically. Please check the data that you have entered and if the source data is correct please state "**REJECTION**" in the description box and expand the error if needed in the "detailed notes" section.

'Query' Discrepancy Notes

A Query Note Type is used to ask a question about data provided in a CRF. For example, you might create a Query Discrepancy Note to check data that seems incomplete or incorrect, or a more general query about what information to input.

Please use the following controlled words in the "description" box to describe your query and use the "detailed notes" section to expand on your query:

i) "**OC**" if there is a query/error relating to OpenClinica.

ii) "**Data**" if the queries relate to how to enter source data into OC.

iii) "**Other**" for all other queries.

It is important to email the COHORT coordinators when entering a "Query" type discrepancy note. These types of Notes begin a conversation thread by you (the originator) assigning the discrepancy note to another user who can answer the question. By selecting 'Email assigned User' the question will be forwarded directly to the users' email account (use cohortcoordination@medschl.cam.ac.uk).

If this is not selected the user can still see that there is a Note that requires their attention by clicking 'Notes and Discrepancies' on the menu bar at the top of the Home page. The page will then display a link that shows specifically which Notes you are assigned. We recommend using the Email check box to ensure your query is answered promptly.

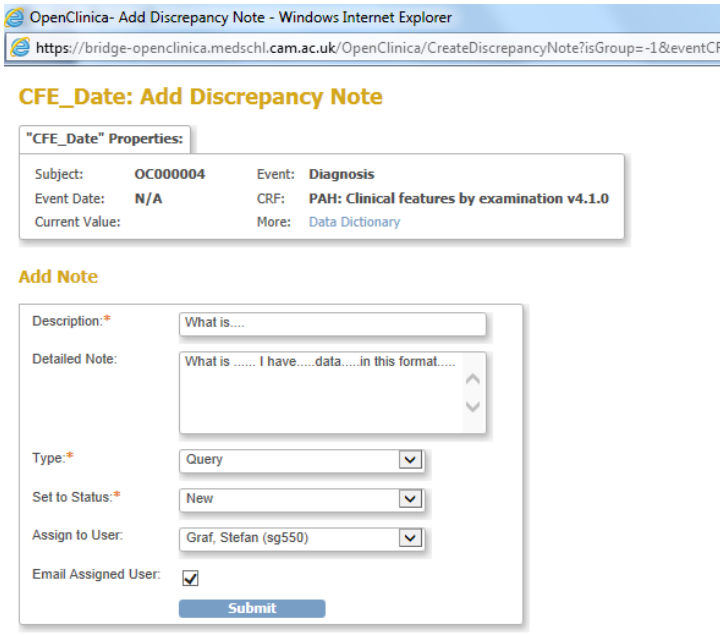
This is a typical workflow:

1. The originator creates the Query Note. The Status is New  . The originator assigns the query to the user who can answer the question.

2. The user it is assigned to updates the Query by adding a new 'child' Discrepancy Note to the parent thread. The user might or might not modify the data value in conjunction with the update. If the user believes the issue is resolved, the user sets the Status to Resolution Proposed  and assigns it to the originator, or, if the user believes further consideration is required, the user sets the Status to Updated  .

3. The originator reviews the response and if satisfied, marks it as Closed  , or adds further comments, assigns it back to the user, and sets the Status to Updated.

Please select appropriate users to answer questions. Bioinformatics staff (Stefan or Matthias) should be assigned where the question is technical or related to the function of the CRF. The research coordinator (Katherine or Carmen) should be assigned where the question is more general. For patient or site specific queries, please assign them to your local PI or your colleague.



Discrepancy Note Status

The status of a note is indicated by the colour of the flag next to the CRF data and in the list under the 'Notes and Discrepancies' section. It provides an indication of who is responsible for the next step.

For **Annotations** the status will always be Not Applicable as no further action is required.

For **Query's and Failed Validation Checks** the initial status should be set to **New**, when responding it should be set to either **Updated** or **Resolution Proposed** (the thread can flick between these two statuses as resolution is discussed in a **Query**), and then finally once dealt with a further note should have the status **Closed** so that no further notes can be added to the thread.

More information

For more detailed information on the use of Discrepancy note please visit the Help Section on OpenClinica:
https://docs.openclinica.com/3.1/openclinica-user-guide/monitor-and-manage-data/notes-and-discrepancies

Appendix 6

Variables recorded in OpenClinica as part of the NIHR BRIDGE PAH Study		
Group	**eCRF**	**Item**
Diagnosis	ID	Subject group (Case / Relative) Diagnosis (idiopathic PAH / heritable PAH / PVOD / PCH) BRIDGE ID Cohort ID
	Demographics	Date of birth Gender Self-declared ethnicity
	Clinical features at diagnosis	Date of diagnosis Date of symptom onset Infection at time of onset? (yes / no) Syncope? (yes / no) Haemoptysis? (yes / no) Raynaud's? (yes / no)
	Clinical features by examination	Date of physical examination Height Weight Body mass index Body surface area Resting oxygen saturations Supplemental oxygen? (yes / no) Amount of supplemental oxygen Systolic blood pressure (BP) Diastolic BP Mean BP Heart rate Jugular venous pressure Digital clubbing? (yes / no) Spider naevi? (yes / no) Peripheral oedema? (yes / no) Ascites? (yes / no)
	Functional class	Functional class
	Haemodynamics	Right heart catheterisation done? (yes / no) Date Supplemental oxygen? (yes / no) Amount of oxygen Heart rate Systolic BP Diastolic BP Mean BP Left ventricular end diastolic pressure

		Pulmonary capillary wedge pressure Systolic pulmonary artery pressure Diastolic pulmonary artery pressure Mean pulmonary artery pressure Right atrial pressure Cardiac output 1 Cardiac output 1 method (thermodilution / direct Fick / indirect Fick) Cardiac output 2 Cardiac output 2 method (thermodilution / direct Fick / indirect Fick) Peripheral arterial oxygen saturation Mixed venous oxygen saturation
		Vasodilator challenge performed? (yes / no) Date Supplemental oxygen? (yes / no) Amount of oxygen Heart rate Vasodilator used Vasodilator responder? (yes / no) Systolic BP Diastolic BP Mean BP Left ventricular end diastolic pressure Pulmonary capillary wedge pressure Systolic pulmonary artery pressure Diastolic pulmonary artery pressure Mean pulmonary artery pressure Right atrial pressure Cardiac output 1 Cardiac output 1 method (thermodilution / direct Fick / indirect Fick) Cardiac output 2 Cardiac output 2 method (thermodilution / direct Fick / indirect Fick) Peripheral arterial oxygen saturation Mixed venous oxygen saturation
	Lung function tests	Lung function testing done? (yes / no) Forced expiratory volume 1 second (FEV_1) FEV_1 (% predicted) Forced vital capacity (FVC) FVC (% predicted) Total lung capacity (TLC) TLC (% predicted) Transfer coefficient (KCO) KCO (% predicted)

		Alveolar volume (VA) VA (% predicted) Sleep study done? (yes / no) Diagnosis (normal / obstructive sleep apnoea / central sleep apnoea) Supplemental oxygen (yes / no) Amount of oxygen Average nocturnal oxygen saturation Desaturation index
	Arterial blood gas (ABG)	ABG done? (yes / no) ABG date Supplemental oxygen (yes / no) Amount of oxygen pH H^+ P_aO_2 P_aCO_2
	Exercise tests	Exercise test done? (yes / no) Type of walk test (corridor / shuttle) Corridor length Date of test Time of test Supplemental oxygen (yes / no) Amount of oxygen Distance walked Peripheral oxygen saturation pre-walk Peripheral oxygen saturation post-walk
		Cardiopulmonary exercise test done? (yes / no) Date Time Height Weight Supplemental oxygen (yes / no) Amount of oxygen FEV_1 FVC Systolic BP rest Diastolic BP rest Mean BP rest O_2 saturation rest O_2 pulse rest VCO_2 at anaerobic threshold VCO_2 nadir $PETCO_2$ at anaerobic threshold VO_2 at anaerobic threshold

		$PETCO_2$ max Systolic BP peak Diastolic BP peak Mean BP peak Heart rate peak Oxygen saturation peak VO_2 peak Peak work rate Peak ventilation Peak RER $PETCO_2$ peak O_2 pulse peak VO_2-WR slope VE-VCO_2 slope Oxygen utilisation efficiency slope
	Clinical blood tests	Autoimmune screen done? Date ANA result (positive / negative) Anti-cardiolipin result (positive / negative) Anti dsDNA result (positive / negative) Anti SCL70 result (positive / negative) Anti-centromere result (positive / negative) Anti Ro result (positive / negative) Anti ENA result (positive / negative) ANCA result (positive / negative)
		Cardiac blood tests done? (yes / no) Date Bone natriuretic peptide (BNP) unit BNP level NT-ProBNP unit NT-ProBNP level Troponin unit Troponin level Urate unit Urate level
		Haematology sample done? (yes / no) Date Haematocrit unit Haematocrit Hb unit Hb Platelet unit Platelet count Red cell distribution width White blood cell count

		Monoclonal band on serum electrophoresis? (paraprotein band yes / no)
		Inflammatory tests done? (yes / no) Date C reactive protein (CRP) unit CRP level High sensitivity CRP unit High sensitivity CRP level
		Iron studies done? (yes / no) Date Ferritin unit Ferritin level Iron unit Iron level Iron binding capacity Transferrin unit Transferrin level
		Lipids tested? (yes / no) Date Cholesterol unit Cholesterol level High density lipoprotein level Low density lipoprotein level Triglycerides level
		Liver function tests done? (yes / no) Date Alkaline phosphatase Albumin Alanine transferase Aspartate transferase Bilirubin Lactate dehydrogenase Total protein
		Renal function checked? (yes / no) Date Sodium Potassium Urea Creatinine unit Creatinine level Estimated glomerular filtration rate (eGFR) unit eGFR level
		Serological tests done? (yes / no) Date Hepatitis B

		Human immunodeficiency virus
		Thyroid function done? (yes / no) Date Free thyroxine (T4) Thyroid stimulating hormone
	Echocardiogram	Echocardiogram done? (yes / no) Date Pericardial effusion present? (yes / no) Tricuspid annular plane systolic excursion Right atrial area Left atrial size Right ventricle morphology (normal / abnormal)
	Electrocardiogram	ECG done? (yes / no) Date Rhythm (sinus / atrial fibrillation / atrial flutter / other) QRS duration Right bundle branch block present? (yes / no) Dominant R wave present? (yes / no)
	Imaging studies	Imaging studies performed? (yes / no) Ventilation – perfusion scan (V/Q) (yes / no) Date Report Pulmonary embolus on VQ? (yes / no)
		CT (yes / no) CTPA (yes / no) Date Report HRCT (yes / no) Date Report Emphysema (no / minimal / mild / moderate / severe) Fibrosis (no / minimal / mild / moderate / severe) CTEPH (yes / no)
		Cardiac MRI (yes / no) Date Report Heart rate Right ventricular ejection fraction Right ventricular diastolic volume Right ventricular systolic volume Left ventricular ejection fraction

		Left ventricular diastolic volume Left ventricular systolic volume
		Abdominal ultrasound (yes / no) Date Report Portal venous flow (normal / abnormal)
Continuous	Risk factors	High altitude exposure? (yes / no) Date of high altitude exposure Duration of high altitude exposure
		Current smoker? (yes / no) Smoking history (pack years)
		Previous pregnancies? (yes / no) Any complications? (yes / no)
		Anorexigen use? (yes / no) Start date End date Duration
		Recreational drug use? (yes / no) Name Start date End date Duration
		Chemotherapy? (yes / no) Name Start date End date Duration
		Other drugs and toxins? (yes / no) Name Start date End date Duration
	Associated conditions	Associated disease present? (yes / no)
		Congenital heart disease present? (yes / no) Eisenmenger syndrome present? (yes / no) Post-tricuspid valve defect? (yes / no) Pre-tricuspid valve defect? (yes / no) Congenital heart disease corrected? (yes / no) Year of surgery Other congenital heart disease (yes / no) International Classification of Diseases (ICD) 10 code
		Connective tissue disease present? (yes / no) Date of diagnosis

		Portal hypertension present? (yes / no) Date of diagnosis
		HIV? (yes / no) Date of diagnosis
		Chronic haemolytic anaemia present? (yes / no) Date of diagnosis
		Schistosomiasis present? (yes / no) Date of diagnosis
		Other associated disease present? ICD 10
	Family history	Family history of PAH (yes / no) Relationship Year of birth Alive (yes / no) Cause of death known?
	Clinical features by history	Comorbidities? (yes / no) Date of diagnosis ICD 10 code
	Drug therapies (not PAH specific)	Other drug therapy? (yes / no) Name Start date End date Unit Dose Regular or as required
	Drug therapies (PAH specific)	Soluble guanylate cyclase antagonists? (yes / no) Name Start date End date Unit Dose Trial drug? Placebo controlled? Duration of placebo-controlled phase
		Prostacyclin analogue (yes / no) Name Start date End date Unit Dose Trial drug? Placebo controlled? Duration of placebo-controlled phase
		Phosphodiesterase 5 inhibitor (yes / no)

		Name Start date End date Unit Dose Trial drug? Placebo controlled? Duration of placebo-controlled phase
		Endothelin receptor antagonist (yes / no) Name Start date End date Unit Dose Trial drug? Placebo controlled? Duration of placebo-controlled phase
		Prostanoids receptor analogue (yes / no) Name Start date End date Unit Dose Trial drug? Placebo controlled? Duration of placebo-controlled phase
		Oral anticoagulant (yes / no) INR dependent? Start date End date Unit Dose
		Adjunctive therapies? Name Start date End date Unit Dose
Suspension	Suspension	Date of suspension Reason for suspension (death / transplantation / lost to follow up / left county / administrative) Cause of death known? (yes / no) Death certificate available? (yes / no)
		Part 1a Part 1b

		Part 1c Part 2

Appendix 7

The R script that I wrote for merging different datasets and creating / calculating new clinical variables (text proceeded by # are annotations explaining the function of the code):

```
#Specify programs and files to use then run checks of the data:
require(xlsx)
options(stringsAsFactors = FALSE)
args <- commandArgs(trailingOnly = TRUE)
bridge_mapping <- ""
cohort_mapping <- ""
normalised_data <- ""
verification_data <- ""
legacy_data <- ""
output_dir <- ""

i <- 1
while(i <= length(args)){
  if (args[i] == "--bridge-mapping"){
    bridge_mapping <- args[i+1]
    i <- i + 2
  } else if (args[i] == "--cohort-mapping"){
    cohort_mapping <- args[i+1]
    i <- i + 2
  } else if (args[i] == "--clean-data"){
    normalised_data <- args[i+1]
    i <- i + 2
  } else if (args[i] == "--verification-data") {
    verification_data <- args[i+1]
    i <- i + 2
  } else if (args[i] == "--paris1") {
    paris1_file <- args[i+1]
    i <- i + 2
  } else if (args[i] == "--paris2") {
    paris2_file <- args[i+1]
    i <- i + 2
  } else if (args[i] == "--diagnosis_original") {
      diagnosis_orginial_file <- args[i+1]
      i <- i + 2
  } else if (args[i] == "--drugs") {
    drug_exp_file <- args[i+1]
    i <- i + 2
  } else if (args[i] == "--legacy-data") {
    legacy_data <- args[i+1]
    i <- i + 2
  } else if (args[i] == "--output-dir") {
    output_dir <- args[i+1]
    i <- i + 2
  #  print(args[i])
  }
}

if (cohort_mapping == "") {
  stop("Please provide cohort mapping")
} else if (!file.exists(cohort_mapping)) {
```

```
  stop("Please check cohort id file exists")
} else if (file.exists(cohort_mapping)){
  cohort_id <- read.table(file=cohort_mapping, na.strings = c("", "NA", "NULL"), stringsAsFactors = F,
encoding = "UTF-8", sep = ("\t"), header = T)
   if (dim(cohort_id)[2] != 3){
     stop("Check structure of cohort id file")
   } else if (dim(cohort_id[duplicated(cohort_id[,2]),])[1] !=0) {
     print(cohort_id[duplicated(cohort_id[,2]) | duplicated(cohort_id[,2], fromLast=T),])
     stop("duplicated OC ID in cohort id file found")
   } else if (dim(cohort_id[duplicated(cohort_id[,3]),])[1] !=0) {
     stop("duplicated COHORT ID in cohort id file found")
   }
}

if (bridge_mapping == "") {
  stop("Please provide BRIDGE mapping")
} else if (!file.exists(bridge_mapping)) {
  stop("Please check BRIDGE ID file exists")
} else if (file.exists(bridge_mapping)){
  bridge_id <- read.table(file=bridge_mapping, na.strings = c("", "NA", "NULL"), stringsAsFactors = F,
encoding = "UTF-8", sep = ("\t"), header = T)
  if (dim(bridge_id)[2] != 3){
   stop("Check structure of BRIDGE ID file")
 } else if (dim(bridge_id[duplicated(bridge_id[,2]),])[1] !=0) {
   print(bridge_id[duplicated(bridge_id[,2]) | duplicated(bridge_id[,2], fromLast=T),])
   stop("duplicated OC ID in BRIDGE ID file found")
 } else if (dim(bridge_id[duplicated(bridge_id[,3]),])[1] !=0) {
   print(bridge_id[duplicated(bridge_id[,3]) | duplicated(bridge_id[,3], fromLast=T),])
   stop("duplicated BRIDGE ID in BRIDGE ID file found")
 }
}

if (normalised_data == "") {
  stop("Please provide normalised data")
} else if (!file.exists(normalised_data)) {
  stop("Please check clean Rdata file exists")
} else if (file.exists(normalised_data)){
  load(file = normalised_data)
}

if (verification_data == "") {
  stop("Please provide verification data")
} else if (!file.exists(verification_data)) {
  stop("Please check verification spreadsheet exists")
} else if (file.exists(verification_data)){
  verification <- read.xlsx(file=verification_data, 1, header = T, stringsAsFactors = F)
}

if (paris1_file == "") {
  stop("Please provide Paris1 data")
} else if (!file.exists(paris1_file)) {
  stop("Please check paris1 spreadsheet exists")
} else if (file.exists(paris1_file)){
  Paris_Fhx1 <- read.xlsx(file=paris1_file, sheetIndex = 1, startRow = 2, header = T, as.data.frame = T)
}

if (paris2_file == "") {
```

```
  stop("Please provide Paris2 data")
} else if (!file.exists(paris2_file)) {
  stop("Please check paris2 spreadsheet exists")
} else if (file.exists(paris2_file)){
  Paris_Fhx2 <- read.xlsx(file=paris2_file, sheetIndex = 1, startRow = 1, header = T, as.data.frame = T)
}

if (diagnosis_orginial_file == "") {
   stop("Please provide Diagnosis original file")
} else if (!file.exists(diagnosis_orginial_file)) {
   stop("Please check diagnosis_original spreadsheet exists")
} else if (file.exists(diagnosis_orginial_file)){
   Diagnosis_original <- read.xlsx(file=diagnosis_orginial_file, sheetIndex = 1, startRow = 1, header = T,
as.data.frame = T)
}

if (drug_exp_file == "") {
  stop("Please provide drug exposure data")
} else if (!file.exists(drug_exp_file)) {
  stop("Please check drug exposure spreadsheet exists")
} else if (file.exists(drug_exp_file)){
  drug_induced_confirmed <- read.csv(file = drug_exp_file, header = T, stringsAsFactors = F, na.strings
= "")
}

if (legacy_data == "") {
  stop("Please provide legacy spreadsheet")
} else if (!file.exists(legacy_data)) {
  stop("Please check legacy spreadsheet exists")
} else if (file.exists(legacy_data)){
  legacy <- read.csv(file=legacy_data, header = T, stringsAsFactors = F, na.strings = c("", "NA"))
}

if (output_dir == "") {
  stop("Please provide Output directory")
} else if (!dir.exists(output_dir)) {
  stop("Please check Output directory exists")
}

###Prepare BRIDGE IDs:
colnames(bridge_id) <- c("Centre","id", "pah_subject_id_bridge_validated")

###Prepare COHORT IDs:
colnames(cohort_id) <- c("Centre","id", "pah_subject_id_cohort_validated")

####
#Merge all ids together:
data.checked <- merge(data.clean, bridge_id, by.x = "id", by.y = "id", all.x=T)
data.checked <- merge(data.checked, cohort_id, by.x = "id", by.y = "id", all.x=T)

#Define functions to correct errors in the data automatically:
replace_NA <- function(id, item) {NA}
replace_date_diag <- function(id, item){
  data.clean[data.clean[,"id"] == id, item]
}
multiply10 <- function(id, item) {
  data.clean[data.clean[,"id"] == id, item] * 10
```

```
}

multiply100 <- function(id, item) {
  data.clean[data.clean[,"id"] == id, item] * 100
}

divide1000 <- function(id, item) {
  data.clean[data.clean[,"id"] == id, item] * 0.001
}

cbt_renal_creatinine_mmolpl <- function(id, item) {
   data.clean[data.clean[,"id"] == id, "abg_paco2"] * 100
}

ver.idx <- which(is.na(verification$check) & !is.na( verification$func))
for (i in ver.idx){
   if (verification[i,"func"]!="replace_value"){
      fun <- get(verification[i,"func"], mode = "function")
      id <- verification[i,"id"]
      item  <- verification[i, "item"]
      data.checked[data.checked[,"id"]==id, item] <- fun(id, item)
   }

   # replace with specific value from the verification spreadsheet:

   if (verification[i,"func"]=="replace_value"){
      id <- verification[i,"id"]
      item  <- verification[i, "item"]
      value <- verification[i, "value"]
      data.checked[data.checked[,"id"]==id, item] <- value
   }
}

#Create new clinical variables:
data.checked$DOB <- as.Date(paste(data.checked$birth_day, data.checked$birth_month,
data.checked$birth_year, sep="/", collapse =NULL), "%d/%m/%Y")
data.checked$hb_date2 <- as.Date(data.checked$hb_date)
data.checked$cfh_date_of_diagnosis <- as.Date(data.checked$cfh_date_of_diagnosis)
data.checked$age_diagnosis <- as.numeric((data.checked$hb_date2 - data.checked$DOB)/365.25)
data.checked$age_diagnosis <- ifelse(is.na(data.checked$age_diagnosis),
as.numeric((data.checked$cfh_date_of_diagnosis - data.checked$DOB)/365.25),
data.checked$age_diagnosis)

data.checked$pcwp_adj <- ifelse(is.na(data.checked$hb_pawp_m) & data.checked$hb_lvedp <=15,
data.checked$hb_lvedp, ifelse((data.checked$hb_pawp_m > 15 & data.checked$hb_lvedp <=15),
data.checked$hb_lvedp, ifelse(data.checked$hb_pawp_m <=15, data.checked$hb_pawp_m, NA)))

data.checked$pvr <- (data.checked$hb_pap_m - data.checked$pcwp_adj) /
data.checked$hb_cardiac_output_value_1
data.checked$ci <- (data.checked$hb_cardiac_output_value_1 / data.checked$bs_bsa)

data.checked$pp <- data.checked$hb_pap_s - data.checked$hb_pap_d
data.checked$ca <- (data.checked$hb_cardiac_output_value_1 / data.checked$hb_heart_rate) /
data.checked$pp
data.checked$rc <- data.checked$pvr * data.checked$ca * 60

#Create variable vasoresponder if meets strict criteria.
```

```
#As some data missing vasoresponder 2 removes need for maintaining CO with NO challenge
data.checked$vasoresponder <- as.factor(ifelse((data.checked$hv_pap_m <40 &
(data.checked$hb_pap_m - data.checked$hv_pap_m) >10 &
(data.checked$hb_cardiac_output_value_1 / data.checked$hv_cardiac_output_value_1) >0.8),
c("vasoresponder"),c("non-responder")))
data.checked$vasoresponder2 <- as.factor(ifelse((data.checked$hv_pap_m <40 &
(data.checked$hb_pap_m - data.checked$hv_pap_m) >10), c("vasoresponder"), c("non-responder")))

#Only use 6mwt performed on 30m corridor for analysis
data.checked$smwd <- ifelse(data.checked$ep_1_corridor_length == 30,
data.checked$ep_1_distance_meters, NA)

#Determine lung function status
data.checked$fev1_fvc <- data.checked$lf_fev1_liters / data.checked$lf_fvc_liters
data.checked$obstructive_pfts <- as.factor(ifelse(data.checked$fev1_fvc < 0.7, c("Obstructive"),
ifelse(data.checked$lf_fvc_pc<80, c("Restrictive"), c("Normal"))))

data.checked$ethnic_category <- ifelse(data.checked$ethnic_category == "A", "British",
 ifelse(data.checked$ethnic_category == "B", "Irish",
  ifelse(data.checked$ethnic_category == "C", "Traveller",
   ifelse(data.checked$ethnic_category == "D", "Other White",
    ifelse(data.checked$ethnic_category == "E", "White & Black Carribean",
     ifelse(data.checked$ethnic_category == "F", "White & Black African",
      ifelse(data.checked$ethnic_category == "G", "White & Asian",
       ifelse(data.checked$ethnic_category == "H", "Any other mixed ethnicity",
        ifelse(data.checked$ethnic_category == "I", "Indian",
         ifelse(data.checked$ethnic_category == "J", "Pakistani",
          ifelse(data.checked$ethnic_category == "K", "Bangladeshi",
           ifelse(data.checked$ethnic_category == "L", "Chinese",
            ifelse(data.checked$ethnic_category == "M", "Other Asian",
             ifelse(data.checked$ethnic_category == "N", "African",
              ifelse(data.checked$ethnic_category == "O", "Carribbean",
               ifelse(data.checked$ethnic_category == "P", "Other Black",
                ifelse(data.checked$ethnic_category == "Q", "Arab",
                 ifelse(data.checked$ethnic_category == "R", "Other",
                  ifelse(data.checked$ethnic_category == "Z", "Not stated", "ERROR"))))))))))))))))))


# identify legacy subjects:
data.checked <- merge(data.checked, legacy, by.x = "pah_subject_id_bridge_validated", by.y =
"legacy_id", all.x = T)
data.checked$legacy2 <- ifelse(is.na(data.checked$legacy), "prosp", "legacy")

# Diagnosis verified:
##PVOD_PCH <- unique(indInfo[which(grepl("PVOD", indInfo$pah_subject_diagnosis) |
##                                 grepl("PCH", indInfo$pah_subject_diagnosis)),"id"])
##data.checked$diagnosis_verified <- ifelse(data.checked$id %in% PVOD_PCH, "PVOD/PCH", "PAH")
##data.checked$diagnosis_verified <- ifelse(data.checked$id %in% Diagnosis_original$ID,
"PVOD/PCH", data.checked$diagnosis_verified)
data.checked$diagnosis_verified <- data.checked$diagnosis
idx <- which(!is.na(data.checked$sub_cause_diagnosis_verified))
data.checked$diagnosis_verified[idx] <- data.checked$sub_cause_diagnosis_verified[idx]
idx <- which(data.checked$sub_cause_diagnosis_verified == "APAH")
data.checked$diagnosis_verified[idx] <- paste(data.checked$sub_cause_diagnosis_verified[idx],
data.checked$sub_cause_apah[idx], sep=":")
```

```
# Family history (yes or no):
relatives_ocid <- unique(indInfo[which(indInfo$pah_subject_id_group_check=="relative"),"id"])

#Take only patients in whom a family history of PAH is certain (ie ignore UNK and no responses in
family_pah)
Confirmed_Fhx <- family_history[which(family_history$family_pah=="yes"),]
Confirmed_Fhx <- Confirmed_Fhx[which(!Confirmed_Fhx$id %in% relatives_ocid),]

#Paris FHx:
Paris_fhx_ids <- c(as.character(Paris_Fhx1[which(Paris_Fhx1$Diagnosis=="Familial PAH"),
"BRIDGE.ID"]),
            as.character(Paris_Fhx2[which(Paris_Fhx2$PAH=="familial"), "BRIDGE"]))

Paris_fhx_ids <- data.checked[which(data.checked$pah_subject_id_bridge_validated %in%
Paris_fhx_ids),
                    c("pah_subject_id_bridge_validated", "id")]

FPAH_list <- c(Confirmed_Fhx$id, Confirmed_Fhx$family_ocid, Paris_fhx_ids$id)
FPAH_list <- FPAH_list[which(!is.na(FPAH_list))]
FPAH_list <- FPAH_list[!duplicated(FPAH_list)]
FPAH_list <- FPAH_list[!FPAH_list %in% relatives_ocid]
data.checked$FHx <- ifelse(data.checked$id %in% FPAH_list, "Yes", "No")

#### Determine drug exposures:
data.checked$drug_exposure <- ifelse(data.checked$id %in% drug_induced_confirmed$id, "Yes",
"No")
data.checked$amphetamines <- ifelse(data.checked$id %in%
drug_induced_confirmed[which(!is.na(drug_induced_confirmed$Rec_drug)),"id"],
                    "Amphetamines", NA)
data.checked$dasatinib <- ifelse(data.checked$id %in%
drug_induced_confirmed[which(!is.na(drug_induced_confirmed$Chemo_drug)),"id"],
                    "Dasatinib", NA)
data.checked$anorexigen <- ifelse(data.checked$id %in%
drug_induced_confirmed[which(!is.na(drug_induced_confirmed$Anorexigen_drug)),"id"],
                    "Anorexigen", NA)

#Reorder columns and clean data.checked for ease of use in analysis stages:
basic_variables <- c("id","pah_subject_id_bridge_validated", "pah_subject_id_cohort_validated",
            "centre", "diagnosis", "diagnosis_verified", "FHx", "drug_exposure", "amphetamines",
            "dasatinib", "anorexigen", "sex", "ethnic_category", "age_diagnosis")

remove_variables <- c(grep("ad_", names(data.checked), value = T), grep("rf_", names(data.checked),
value = T))
other <- names(data.checked)[!names(data.checked) %in% c(basic_variables, remove_variables)]

data.checked <- data.checked[, c(basic_variables, other)]

print(paste0("Saving to ", output_dir, "/data_checked.RData"))
save(data.checked, file = paste0(output_dir, "/data_checked.RData"))
```

Appendix 8

Tests applied to check the phenotype data			
Type	**eCRF**	**Check**	**Comment**
Errors	All	S_aO_2 >100%	
		Systolic BP > diastolic BP	
		Zero values entered	Sometimes entered where data missing
	Haemodynamics	Date of right heart catheter more than 6 weeks from date of diagnosis	
		Systolic PAP > mean PAP > diastolic PAP > PCWP	
		PCWP > 15 mmHg and no LVEDP available or LVEDP also > 15 mmHg	
		mPAP < 25 mmHg	
	Exercise performance	Corridor length (10, 15, 25 or 30 m in length)	Corridor length and walk test distance mixed up
	Imaging investigations	End systolic volumes > end diastolic volumes	
	Clinical blood test	Hb < 50 g/l	May indicate incorrect units used
	Demographics	Self-declared gender same as WGS inferred gender	
Relationships	Haemodynamics	mPAP – sPAP	
		mPAP – dPAP	
		sPAP – dPAP	
		FEV_1 – FEV_1 % predicted	
		FVC – FVC % predicted	
		TLC – TLC % predicted	
		KCO – KCO % predicted	
		VA – VA % predicted	
	Exercise performance	6mwt distance to functional class	

Appendix 9

Customised proforma used to record radiographic features from computed tomography imaging studies from the time of diagnosis	
Parameter	**Response**
ID	
Date of birth	
Unenhanced CT	Y / N
CTPA	Y / N
HRCT	Y / N
Expiratory CT	Y / N
Pulmonary artery diameter (cm)	
Aorta diameter (cm)	
Ground glass opacification centrilobular pattern DENSITY	None / Subtle / Present
Ground glass centrilobular pattern EXTENT	0 %, < 5 %, 5 % – 25 %, 25 % – 50 %, > 50 %
Ground glass DISTRIBUTION	central (C)/peripheral (P)/zonal (Z) or diffuse (D)
Non-specific mosaic pattern / groundglass opacification	
Neovascularity vessels	Y / N
Arterio-venous malformations	Y / N
Bronchial arteries	Y / N
Largest bronchial artery size	
Interlobular septal thickening	None, Subtle, Present
Mediastinal lymphadenopathy	Y / N
Emphysema	Y / N and % of parenchyma involved
Fibrosis	Y / N and % of parenchyma involved
Pleural effusion	Y / N
Air trapping	Y / N
Comments	
Likely diagnosis	Any suspicion of PVOD or PCH / PAH
CT – computer tomography, CTPA – computer tomography pulmonary angiogram, HRCT – high resolution CT	

Appendix 10

The R scripts I wrote to analyse the data and create the tables that are presented in the Thesis are provided below:

To create the summary tables presented in the Thesis:

1. Specify the raw data being used (*DF*) and specify the clinical variable being assessed (*VARIABLE*; e.g. gender):

```
DF <- clean.data
VARIABLE <- "sex"
```

2. Specify the number of levels in the clinical variable being assessed (*GROUP*; e.g. male/female):

```
n_levels <- length(levels((as.factor(DF[,VARIABLE]))))
numeric_summary <- data.frame(matrix(ncol =  2+3*n_levels, nrow = 0), stringsAsFactors = F)
```

3. In each *GROUP* calculate the median (*med*), 1st and 3rd quartiles (*iqr1* and *iqr2*) for all numeric variables (test_*var;* e.g. age at diagnosis) in the dataframe (*DF_NUMS*) sequentially and add to summary table (*numeric_summary_temp*). Rounded each result to one decimal place.

```
numeric_summary <- data.frame(matrix(ncol =  2+3*n_levels, nrow = 0), stringsAsFactors = F)

for (i in test_var) {
  numeric_summary_temp <- data.frame(matrix(nrow = 1, ncol = 0), stringsAsFactors = F)
  for (t in GROUP){
    temp <- data.frame(variable=i,
      mean = sprintf("%.1f", round(median(DF_NUMS[which(DF_NUMS[,VARIABLE]==t), i], na.rm = T), digits = 2)),
      iqr1 = sprintf("%.1f", round(quantile(DF_NUMS[which(DF_NUMS[,VARIABLE]==t), i], na.rm = T)[2], digits = 2)),
      iqr2 = sprintf("%.1f", round(quantile(DF_NUMS[which(DF_NUMS[,VARIABLE]==t), i], na.rm = T)[4], digits = 2)), stringsAsFactors = F)

    temp[,t] <- paste0(temp$mean, " [", temp$iqr1, " - ", temp$iqr2, "]")
    temp <- temp[5]
    numeric_summary_temp <- cbind(numeric_summary_temp, temp)
  }
```

3. Hypothesis testing based on number (*n_levels*) of groups being compared (2 groups – unpaired Wilcoxon rank sum test, 3 of more groups – Kruskal-Wallis test):

```
  if (n_levels == 2){
    numeric_summary_temp$p_val <- wilcox.test(DF_NUMS[,i] ~ as.factor(DF_NUMS[,VARIABLE]), na.action = na.exclude)$p.value
  }

  if (n_levels>2) {
```

```
   numeric_summary_temp$p_val <- kruskal.test(DF_NUMS[,i] ~
as.factor(DF_NUMS[,VARIABLE]), na.action = na.exclude)$p.value
  }

  numeric_summary_temp$variable <- i
  numeric_summary_temp <- numeric_summary_temp[,c(ncol(numeric_summary_temp),
1:(ncol(numeric_summary_temp)-1))]
  numeric_summary <- rbind(numeric_summary, numeric_summary_temp)
}
```

4. For all categorical variables (*cat_var*) calculate the number and percentage of the total in each group being assessed (*GROUP*):

```
categorical_summary <- data.frame(matrix(nrow = 1, ncol = 0), stringsAsFactors = F)

for(var in cat_var){
  table <- table(DF_CAT[,VARIABLE], DF_CAT[,var])
  temp2 <- data.frame(matrix(nrow = 1, ncol = 0), stringsAsFactors = F)
  temp2[,"variable"] = var
  temp2[,"level_reported"] <- character()

  if (ncol(table) <=2){
    temp2[,"level_reported"] = colnames(table)[1]
  }
  if (ncol(table) >2){
    for (n in 1:length(colnames(table)))
      temp2$level_reported = paste0(temp2$level_reported, " / ", colnames(table)[n])
      temp2$level_reported <- gsub("NA / ", "", temp2[,"level_reported"])
  }

 for (g in GROUP){
    temp <- data.frame(matrix(nrow = 1, ncol = 0), stringsAsFactors = F)
    if (ncol(table) <=2){
      temp[,g] = paste0(table[rownames(table)==g, 1], " [", sprintf("%.1f",
round((prop.table(table,1)[rownames(table)==g, 1]) *100 , digits = 1)), "%]")
      temp2 <- cbind(temp2, temp)
    }
    if (ncol(table) > 2){
      temp[,g] <- NA
      for (n in 1:ncol(table)){
        temp[,g] = paste0((temp[,g]), " / ", table[rownames(table)==g, n], " [", sprintf("%.1f",
round((prop.table(table,1)[rownames(table)==g, n]) *100 , digits = 1)), "%]")
        temp[,g] <- gsub("NA / ", "", temp[,g])
      }
      temp2 <- cbind(temp2, temp)
      }
    }
```

5. Hypothesis testing using Fisher's exact test for non-ordinal variables and the chi square test for ordinal variables:

```
 if (ncol(table) >1 & VARIABLE_ORDINAL=="yes"){
   temp2$p_val <- sprintf("%.3f", round(pvalue(chisq_test(table(DF_CAT[,VARIABLE],
DF_CAT[,var], dnn=c("gt", var)), scores = list("gt" = seq(1:n_levels)))), digits = 3))
 }
 if (ncol(table) >1 & VARIABLE_ORDINAL=="no"){
   temp2$p_val <- sprintf("%.3f", round(fisher.test(table(DF_CAT[,VARIABLE], DF_CAT[,var]),
workspace = 2e9)$p.value, digits = 3))
 }
 if (ncol(table) <=1){
    temp2$p_val <- NA
 }
 categorical_summary <- rbind(categorical_summary, temp2)
}
```

Assessing associations with haemodynamic variables with rank regression models:

1. Select haemodynamic variables and assess their distribution using the Shapiro-Wilk test.

```
test_var <- c("hb_pap_m", "pcwp_adj", "pvr", "ci", "hb_sv_o2", "hb_rap_m")
for (i in test_var){
 print("*******")
 print(i)
 print(shapiro.test((clean.data[,i])))
 print("*******")
}
```

2. Create univariate rank regression models:

```
library(Rfit)
for (i in test_var){
 print("*******")
 print(i)
 print(summary(rfit(get(i) ~ age_diagnosis, clean.data)))
 print("*******")
}

for (i in test_var){
 print("*******")
 print(i)
 print(summary(rfit(get(i) ~ sex, clean.data)))
 print("*******")
}

for (i in test_var){
 print("*******")
 print(i)
 print(summary(rfit(get(i) ~ centre, clean.data)))
 print("*******")
```

```
}
```

3. Create multivariate rank regression models using age, gender and centre as covariates:

```
MV_RHC_model <- data.frame(var=character(), var2 = character(), b=character(),
se=character(), p=character())

for (i in test_var){
  model <- summary(rfit(get(i) ~ (sex + age_diagnosis + centre), clean.data))
  temp <- data.frame(var=character(), var2 = character(), b=character(), se=character(),
p=character())

  for (t in c("sexmale", "age_diagnosis")){
    temp2 <- data.frame(var = i,
      var2 = t,
      b = round(model$coefficients[t,"Estimate"], digits = 3),
      se = round(model$coefficients[t,"Std. Error"], digits = 3),
      p = round(model$coefficients[t,"p.value"], digits = 3))
    temp <- rbind(temp2, temp)
  }

  MV_RHC_model <- rbind(MV_RHC_model, temp)
  MV_RHC_model <- MV_RHC_model[which(MV_RHC_model$p <0.05), ]
}
```

Perform a non-truncated univariate survival analysis:

1. Sequentially assess all variables (*test_var*) in a dataset containing survival data (*survival.data*). The analysis censors patients who are not dead (right censoring). Extract hazard ratio and confidence value for each variable:

```
survival_univariate_no_trunc <- data.frame(variable=character(),
group=character(),HR=character(), CI=character(), p=character(), n=character(),
n_events=character(), stringsAsFactors = F)

for (var in test_var){
  temp <- coxph(Surv(time = survival.data$sur_time2, survival.data$CAUSE, type = "right") ~
(survival.data[,var]))

  temp_df <- data.frame(variable=var,
    HR = round(summary(temp)$conf.int[1,"exp(coef)"], digits = 3),
    group = "",
    CI = paste0(round(summary(temp)$conf.int[1,"lower .95"],digits = 3), " - ",
round(summary(temp)$conf.int[1,"upper .95"], digits = 3)),
    p = round(summary(temp)$coefficients[1,"Pr(>|z|)"], digits = 5),
    n = summary(temp)$n,
    n_events = summary(temp)$nevent, stringsAsFactors = F)
  survival_univariate_no_trunc <- rbind(survival_univariate_no_trunc, temp_df)
}

#more than 2 levels categorical variables:
```

```
survival_univariate2_no_trunc <- data.frame(variable=character(), group=character(),
HR=character(), CI=character(), p=character(), n=character(), n_events=character(),
stringsAsFactors = F)
for (var in three_level_cat){
  temp <- coxph(Surv(time = survival.data$sur_time2, survival.data$CAUSE, type = "right") ~
(survival.data[,var]))

   for( n in 1:(nrow(summary(temp)$coefficients))){
     temp_df <- data.frame(variable=var,
       group= row.names(summary(temp)$coefficients)[n],
       HR = round(summary(temp)$conf.int[n,"exp(coef)"], digits = 3),
       CI = paste0(round(summary(temp)$conf.int[n,"lower .95"],digits = 3), " - ",
round(summary(temp)$conf.int[n,"upper .95"], digits = 3)),
       p = round(summary(temp)$coefficients[n,"Pr(>|z|)"], digits = 5),
       n = summary(temp)$n,
       n_events = summary(temp)$nevent, stringsAsFactors = F)

     survival_univariate2_no_trunc <- rbind(survival_univariate2_no_trunc, temp_df)
  }
}

survival_univariate_all_no_trunc <- rbind(survival_univariate_no_trunc,
survival_univariate2_no_trunc)
```

Left truncated survival analyses:

1. The same R code as the non-truncated analysis above was used except replacing the survival function with the line below:

```
temp <- coxph(Surv(time = survival.data$L_TRUNC, time2 = survival.data $sur_time2,
survival.data $CAUSE, type = "counting") ~ (survival.data [,var]))
```

Appendix 11

Data completion rates for diagnostic data captured in OpenClinica		
Item name	**Number complete**	**% complete**
Day of birth	1065	99.5
Month of birth	1065	99.5
Year of birth	1065	99.5
Centre	1065	99.5
Diagnosis group	1065	99.5
OpenClinica ID	1065	99.5
Gender	1065	99.5
Diagnosis	1064	99.4
BRIDGE ID	1061	99.2
Date of diagnosis	1040	97.2
mPAP (mmHg)	1003	93.7
CO (L/min)	957	89.4
Functional class	940	87.9
Day of visit	938	87.7
Month of visit	938	87.7
Year of visit	938	87.7
Ethnic category	937	87.6
Haemoptysis at onset	923	86.3
Raynaud's syndrome	923	86.3
Infection at onset	922	86.2
Syncope at onset	922	86.2
Symptom onset known	917	85.7
PVR (dynes)	898	83.9
Date of RHC	895	83.6
RAP (mmHg)	876	81.9
CO method 1	857	80.1
RHC supplemental oxygen used	845	79.0
6mwt distance (m)	834	77.9
PCWP (mmHg)	816	76.3
Date of walk test	799	74.7
sPAP (mmHg)	779	72.8
dPAP (mmHg)	776	72.5
Ankle swelling	775	72.4
Ascites	775	72.4
Clubbing	774	72.3
Spider naevi	774	72.3
Type of walk test	773	72.2

Date of examination	764	71.4
Weight (kg)	755	70.6
Walk test supplemental O_2 used	742	69.3
Height (cm)	736	68.8
Hb unit	734	68.6
Hb (g/l)	726	67.9
Serum electrophoresis	727	67.9
Sodium (mmol/l)	726	67.9
WBC ($x10^9$/l)	722	67.5
Creatinine unit	722	67.5
Platelets ($x10^9$/l)	720	67.3
Potassium (mmol/l)	720	67.3
eGFR unit	719	67.2
Urea (mmol/l)	716	66.9
HCT unit	715	66.8
Creatinine (μmol/l)	715	66.8
Systemic diastolic BP (mmHg)	711	66.4
Systemic systolic BP (mmHg)	711	66.4
ALP (iu)	703	65.7
Examination supplemental O_2 use	703	65.7
Bilirubin (μmol/l)	699	65.3
Date of echocardiogram	697	65.1
RV on echocardiogram	693	64.8
ALT (iu/l)	692	64.7
Albumin (g/l)	689	64.4
Resting S_aO_2 (%)	664	62.1
RHC S_aO_2 (%)	662	61.9
BSA	660	61.7
Pre-walk S_aO_2 (%)	660	61.7
S_vO_2 (%)	657	61.4
FEV_1 (l)	651	60.8
Date of lung function tests	648	60.6
FVC (l)	647	60.5
FEV_1 (% predicted)	641	59.9
BMI	637	59.5
Date of height and weight	637	59.5
Dominant R wave	626	58.5
RBBB	626	58.5
FVC (% predicted)	626	58.5
Pericardial effusion	623	58.2
Date of full blood count	607	56.7
Date of renal function	606	56.6

Date of ECG	604	56.4
ECG done	603	56.4
Heart rhythm	604	56.4
HR (bpm)	598	55.9
Date of liver function tests	593	55.4
Post walk S_aO_2 (%)	591	55.2
Anti-Centromere Ab	574	53.6
Anti-Rho Ab	573	53.6
KCO (mmol/min/kPa/l)	573	53.6
Anti-ENA Ab	572	53.5
Anti-Scl70 Ab	572	53.5
ANA	571	53.4
ANCA	568	53.1
Total protein (g/l)	567	53.0
HCT	559	52.2
Anti dsDNA Ab	553	51.7
RHC mean BP (mmHg)	551	51.5
RHC systolic BP (mmHg)	549	51.3
Anti-cardiolipin Ab	548	51.2
RHC diastolic BP (mmHg)	548	51.2
KCO (% predicted)	544	50.8
TSH (mu/l)	537	50.2
CRP (mg/l)	527	49.3
CT CTEPH	528	49.3
CT fibrosis	519	48.5
CT emphysema	517	48.3
Urate unit	496	46.4
Date of symptoms onset	496	46.4
VA	496	46.4
RHC HR (bpm)	495	46.3
BNP unit	494	46.2
NT-ProBNP unit	492	46.0
Troponin unit	488	45.6
CTPA report	483	45.1
Date of autoantibody screen	474	44.3
Corridor length (m)	471	44.0
TLC	469	43.8
TLC (% predicted)	456	42.6
V/Q scan report	454	42.4
Date of thyroid function tests	449	42.0
QRS axis	442	41.3
PE on V/Q scan	440	41.1

CI (L/min/m^2)	436	40.7
Date of inflammatory markers	430	40.2
Date of cardiac blood tests	404	37.8
VA (% predicted)	393	36.7
AST (iu/l)	390	36.4
Date of CTPA	390	36.4
Date of vasodilator study	370	34.6
Ferritin unit	366	34.2
FT4 (pmol/l)	366	34.2
Transferrin unit	364	34.0
eGFR (ml/min)	358	33.5
Vasodilator response	353	33.0
Hepatitis B serology	352	32.9
HIV serology	352	32.9
ABG acidity measurement	349	32.6
$PaCO_2$ (kPa)	349	32.6
PaO_2 (kPa)	349	32.6
pH	348	32.5
Vasodilator used	345	32.2
Iron (μmol/l)	338	31.6
Date of V/Q scan	338	31.6
COHORT ID	332	31.0
TAPSE (cm)	321	30.0
ABG supplemental O_2 use	319	29.8
RDW	303	28.3
Ferritin (μg/l)	300	28.0
JVP (cm)	288	26.9
Vasodilator study supplemental oxygen used	285	26.6
MRI report	266	24.9
Date of serological assessment	263	24.6
Vasodilator study mPAP (mmHg)	260	24.3
Date of iron studies	251	23.5
ABG date	248	23.2
Ultrasound report	247	23.1
Vasodilator study sPAP (mmHg)	241	22.5
Vasodilator study dPAP (mmHg)	240	22.4
US portal venous flow	239	22.3
HRCT report	237	22.1
Urate (mmol/l)	230	21.5
Vasodilator study cardiac output method 1	227	21.2
NT-ProBNP (ng/l)	221	20.7
Date of MRI	221	20.7

Total cholesterol (mmol/l)	217	20.3
LVEDP (mmHg)	208	19.4
RA area (cm^2)	204	19.1
Vasodilator study CO (L/min)	204	19.1
Date of HRCT	200	18.7
LDH (iu/l)	187	17.5
Iron binding capacity (μmol/l)	186	17.4
Lung function supplemental oxygen used	179	16.7
Vasodilator study systolic BP (mmHg)	178	16.6
Vasodilator study diastolic BP (mmHg)	176	16.4
MRI RVEF (%)	175	16.4
Vasodilator study mean BP (mmHg)	173	16.2
Vasodilator study S_aO_2 (%)	172	16.1
MRI LVEF (%)	170	15.9
MRI RV diastolic volume (ml)	166	15.5
MRI RV systolic volume (ml)	165	15.4
Troponin (μg/l)	164	15.3
MRI LV diastolic volume (ml)	163	15.2
MRI LV systolic volume (ml)	163	15.2
Triglycerides (mmol/l)	162	15.1
LA size (cm)	162	15.1
Transferrin (g/l)	161	15.0
OSS S_aO_2 (%)	154	14.4
CO method 2	147	13.7
CO value 2 (L/min)	147	13.7
BNP (ng/l)	146	13.6
Time of walk test	145	13.6
Date of abdominal ultrasound	141	13.2
HDL cholesterol (mmol/l)	140	13.1
Vasodilator study PCWP (mmHg)	136	12.7
MRI HR (bpm)	134	12.5
Date of lipids	130	12.1
Systemic mean BP (mmHg)	128	12.0
Vasodilator study PVR (WU)	127	11.9
Vasodilator study S_vO_2 (%)	120	11.2
LDL cholesterol (mmol/l)	112	10.5
Vasodilator study RAP (mmHg)	98	9.2
Sleep study	94	8.8
FEV_1 (z score)	90	8.4
CPET date	84	7.9
CPET supplemental O_2 use	84	7.9
FVC (z score)	82	7.7

K_{CO} (z score)	81	7.6
TLC (z score)	80	7.5
CPET resting S_aO_2 (%)	78	7.3
CPET peak O_2 pulse	77	7.2
CPET peak RER	73	6.8
CPET peak work (W)	73	6.8
CPET height (m)	71	6.6
CPET weight (kg)	71	6.6
CPET peak Ventilation	70	6.5
Examination amount of supplemental O_2 use	64	6.0
CPET resting systolic BP (mmHg)	62	5.8
CPET resting diastolic BP (mmHg)	61	5.7
Vasodilator study CI (L/min/m^2)	61	5.7
CPET FEV_1 (% predicted)	60	5.6
CPET VC (% predicted)	60	5.6
CPET resting O_2 pulse	58	5.4
VA (z score)	58	5.4
Vasodilator study amount of supplemental oxygen used	57	5.3
CI value 2 (L/min/m^2)	53	5.0
RHC amount of supplemental oxygen used	51	4.8
CPET PET CO_2 peak	50	4.7
CPET peak $P_{ET}CO_2$	49	4.6
CPET VE - VCO_2 slope	45	4.2
Vasodilator study cardiac output method 2	38	3.6
Vasodilator study CO value 2 (L/min)	39	3.6
CPET AT VO_2	35	3.3
CPET VCO_2 at AT units	35	3.3
CPET VO_2 - WR slope	35	3.3
ODI	35	3.3
CPET OUES	31	2.9
Walk test amount of supplemental O_2 used	27	2.5
ABG amount of supplemental O_2	24	2.2
hsCRP (mg/l)	22	2.1
Vasodilator study LVEDP (mmHg)	22	2.1
Lung function amount of supplemental oxygen used	23	2.1
CPET $EqCO_2$ at AT	20	1.9
Time of CPET	11	1.0
Vasodilator study CI value 2 (L/min/m^2)	11	1.0
CPET mean BP at rest	10	0.9
Vasodilator used 2	3	0.3
CPET VCO_2 at nadir	2	0.2
CPET supplemental O_2 amount	2	0.2

6mwt – six minute walk test, Ab – antibody, ABG – arterial blood gas, ALP – alkaline phosphatase, ALT – alanine transferase, ANA – anti-nuclear antibody, ANCA - anti-neutrophil cytoplasmic antibody, anti-ENA – anti-extractable nuclear antigen antibody, AST – aspartate transaminase, AT – anaerobic threshold, BMI – body mass index, BNP – bone natriuretic peptide, BP – blood pressure, BSA – body surface area, CI – cardiac index, CO – cardiac output, CO_2 – carbon dioxide, CPET – cardiopulmonary exercise test, CRP – C-reactive protein, CT – computer tomography, CTEPH – chronic thromboembolic pulmonary hypertension, CTPA – computer tomography pulmonary angiography, dsDNA – double stranded DNA, ECG – electrocardiogram, $EqCO_2$ – ventilatory equivalent for carbon dioxide, FEV_1 – forced expiratory volume in 1 second, FT4 – free thyroxine, FVC – forced vital capacity, Hb – haemoglobin, HCT – haematocrit, HDL – high density lipoprotein, HR – heart rate, HRCT – high resolution computer tomography, hsCRP – high sensitivity C-reactive protein, dPAP – diastolic pulmonary artery pressure, eGFR – estimated glomerular filtration rate, JVP – jugular venous pressure, K_{CO} – transfer coefficient for carbon monoxide, LA – left atrium, LV – left ventricle, LVEF – left ventricular ejection fraction, MRI – magnetic resonance imaging, NT-ProBNP – N terminal Pro brain natriuretic peptide, O_2 – oxygen, ODI – oxygen desaturation index, OSS – overnight sleep study, OUES – oxygen utilisation efficiency slope, PCWP – pulmonary capillary wedge pressure, PE – pulmonary embolus, P_{ET} – end tidal partial pressure, PVR – pulmonary vascular resistance, RA – right atrium, RAP – right atrial pressure, RBBB – right bundle branch block, RDW – red blood cell distribution width, RER – respiratory exchange ratio, RHC – right heart catheterisation, RV – right ventricle, RVEF – right ventricular ejection fraction, S_aO_2 – peripheral arterial oxygen saturation, sPAP – systolic pulmonary artery pressure, S_vO_2 – mixed venous oxygen saturation, TAPSE – tricuspid annular plane systolic excursion, TLC – total lung capacity, TSH – thyroid stimulating hormone, US – ultrasound, VC – vital capacity, VCO_2 – carbon dioxide output, VE/VCO_2 – ventilatory equivalents for carbon dioxide, VO_2 – oxygen uptake, WBC – white blood cells, WR – work rate

Appendix 12

Phenotypic differences between patients recruited from different centres

Variable name	Glasgow	Great Ormond Street	Imperial and Hammersmith	Newcastle Freeman	Papworth	Royal Brompton	Royal Free	Royal United Hospital Bath	Sheffield	VU University Medical Center Amsterdam	p corrected
n [%]	79 [10.5%]	8 [1.1%]	233 [30.9%]	54 [7.2%]	114 [15.1%]	53 [7.0%]	29 [3.9%]	23 [3.1%]	116 [15.4%]	44 [5.8%]	
WGS population: African / East-Asian / European / South-Asian (n [%])	0 [0.0%] / 0 [0.0%] / 79 [100.0%] / 0 [0.0%]	0 [0.0%] / 0 [0.0%] / 6 [75.0%] / 2 [25.0%]	11 [4.7%] / 7 [3.0%] / 188 [80.7%] / 27 [11.6%]	0 [0.0%] / 0 [0.0%] / 53 [98.1%] / 1 [1.9%]	1 [0.9%] / 0 [0.0%] / 107 [93.9%] / 6 [5.3%]	0 [0.0%] / 1 [1.9%] / 44 [83.0%] / 8 [15.1%]	2 [6.9%] / 0 [0.0%] / 22 [75.9%] / 5 [17.2%]	0 [0.0%] / 1 [4.3%] / 22 [95.7%] / 0 [0.0%]	1 [0.9%] / 0 [0.0%] / 109 [94.0%] / 6 [5.2%]	2 [4.5%] / 3 [6.8%] / 38 [86.4%] / 1 [2.3%]	<0.001
Age at diagnosis (years)	52.3 [41.1 - 65.8]	7.8 [5.3 - 9.9]	47.5 [35.4 - 65.2]	59.1 [41.1 - 68.7]	47.8 [36.3 - 62.2]	42.9 [33.1 - 53.7]	46.7 [34.2 - 56.9]	62.0 [47.9 - 72.1]	51.3 [34.5 - 64.6]	45.3 [29.0 - 51.3]	<0.001
mPAP (mmHg)	52.0 [44.5 - 59.0]	29.5 [27.5 - 33.0]	54.0 [45.5 - 62.0]	48.0 [39.0 - 56.0]	53.0 [46.0 - 60.0]	53.0 [42.5 - 63.5]	54.0 [45.0 - 64.0]	57.0 [49.5 - 63.5]	55.0 [49.0 - 64.0]	50.0 [43.0 - 59.0]	<0.001
PCWP (mmHg)	8.0 [5.0 - 10.0]	7.5 [7.0 - 8.8]	10.0 [8.0 - 13.0]	8.0 [5.8 - 10.2]	10.0 [8.0 - 11.0]	7.5 [6.0 - 11.0]	11.0 [9.0 - 12.0]	14.0 [10.5 - 15.0]	10.0 [8.0 - 13.0]	8.0 [7.0 - 10.0]	<0.001
CO (L/min)	3.8 [3.0 - 4.7]	3.1 [1.6 - 3.4]	3.7 [2.9 - 4.8]	4.0 [3.3 - 4.7]	3.5 [2.8 - 4.5]	3.9 [2.9 - 5.1]	4.2 [3.3 - 5.7]	3.9 [3.3 - 4.5]	4.0 [3.3 - 5.1]	4.6 [3.8 - 6.0]	0.002
CI (L/min/m²)	2.1 [1.8 - 2.5]	3.0 [2.3 - 3.7]	2.0 [1.6 - 2.6]	2.1 [1.8 - 2.6]	2.0 [1.6 - 2.4]	2.0 [1.6 - 2.8]	2.4 [1.9 - 2.8]	2.0 [1.7 - 2.4]	2.2 [1.8 - 2.8]	2.4 [2.0 - 3.2]	0.005
Vasoresponder: (n [%])	7 [25.9%]	0 [0.0%]	2 [3.6%]	2 [11.8%]	7 [38.9%]	0 [0.0%]	1 [5.9%]	0 [0.0%]	11 [21.6%]	2 [8.3%]	0.004
Weight (kg)	78.4 [64.5 - 94.4]	21.6 [18.5 - 32.3]	73.0 [62.1 - 86.8]	75.0 [62.5 - 89.2]	74.1 [64.4 - 84.0]	77.5 [61.2 - 88.2]	73.0 [63.0 - 84.0]	87.0 [77.7 - 95.5]	78.5 [65.9 - 90.6]	72.0 [63.0 - 82.8]	<0.001
BMI	28.7 [23.6 - 34.7]	16.2 [14.9 - 17.8]	27.6 [23.9 - 32.0]	28.3 [24.0 - 32.6]	27.1 [23.7 - 31.3]	26.8 [22.8 - 33.1]	27.1 [22.1 - 30.0]	31.6 [27.5 - 34.6]	27.9 [25.2 - 32.3]	24.3 [22.3 - 28.7]	<0.001
ALP (iu)	103.0 [78.0 - 178.0]	162.5 [159.0 - 203.5]	82.0 [65.0 - 107.0]	88.5 [77.2 - 105.0]	97.5 [74.0 - 127.5]	69.5 [57.5 - 98.0]	73.0 [59.8 - 90.0]	83.0 [69.0 - 96.5]	84.0 [66.0 - 104.0]	81.0 [62.5 - 105.0]	<0.001
Albumin (g/l)	43.0 [39.0 - 45.0]	36.5 [34.5 - 37.0]	39.0 [36.0 - 42.0]	43.5 [40.8 - 45.0]	40.0 [37.8 - 42.0]	41.0 [38.0 - 44.0]	44.0 [42.0 - 46.0]	44.0 [40.0 - 45.5]	43.0 [40.0 - 45.0]	38.5 [36.0 - 41.8]	<0.001
Sodium (mmol/l)	140.0 [139.0 - 142.0]	141.0 [139.5 - 142.5]	139.0 [137.0 - 141.0]	140.0 [138.0 - 141.0]	139.0 [137.0 - 141.0]	138.0 [136.0 - 139.0]	141.5 [140.0 - 143.0]	138.0 [137.5 - 140.0]	139.0 [138.0 - 141.0]	140.0 [138.0 - 141.0]	<0.001
Creatinine (µmol/l)	88.5 [78.0 - 106.0]	36.0 [35.5 - 52.0]	88.0 [74.5 - 104.5]	95.0 [79.0 - 106.0]	92.0 [82.5 - 106.5]	80.0 [70.0 - 89.0]	79.5 [69.5 - 107.0]	91.0 [69.5 - 102.0]	84.0 [71.0 - 98.2]	87.0 [72.0 - 94.0]	<0.001

Free T4 (pmol/l)	17.5 [14.0 - 20.0]	19.6 [17.4 - 21.0]	15.1 [13.3 - 17.2]	14.4 [13.4 - 17.2]	16.3 [15.8 - 18.9]	13.2 [12.0 - 14.3]	15.8 [14.2 - 17.1]	15.6 [15.2 - 16.0]	16.7 [14.7 - 17.9]	16.6 [14.4 - 18.7]	<0.001
Corridor length (m)	30.0 [30.0 - 30.0]	0.0 [0.0 - 0.0]	30.0 [30.0 - 30.0]	30.0 [30.0 - 30.0]	10.0 [10.0 - 10.0]	15.0 [10.0 - 15.0]	30.0 [30.0 - 30.0]	20.0 [20.0 - 20.0]	15.0 [15.0 - 15.0]	50.0 [30.0 - 50.0]	<0.001
6mwt distance (m)	320.0 [181.0 - 395.5]	240.0 [240.0 - 240.0]	280.0 [120.0 - 405.0]	313.0 [210.0 - 372.8]	343.0 [242.5 - 381.0]	315.0 [250.0 - 400.0]	352.0 [299.0 - 395.0]	209.0 [153.5 - 377.0]	320.0 [180.0 - 412.5]	475.0 [387.0 - 553.0]	<0.001
Post walk S_aO_2 (%)	87.0 [81.2 - 94.0]	97.0 [97.0 - 97.0]	93.0 [86.0 - 97.0]	89.0 [85.0 - 95.0]	91.0 [85.0 - 95.0]	92.0 [86.8 - 97.0]	94.5 [92.2 - 96.8]	92.0 [84.2 - 94.0]	86.0 [80.5 - 90.0]	93.0 [88.5 - 95.0]	<0.001
dPAP (mmHg)	32.0 [26.0 - 36.0]	16.5 [13.5 - 22.5]	33.0 [26.0 - 40.0]	28.0 [22.0 - 34.0]	35.0 [29.2 - 40.0]	35.0 [28.0 - 42.5]	35.0 [28.5 - 46.0]	38.0 [34.0 - 42.0]	36.0 [30.0 - 41.0]	33.0 [27.5 - 39.5]	<0.001
RAP (mmHg)	8.0 [5.0 - 10.0]	5.0 [5.0 - 6.0]	10.0 [6.0 - 13.0]	8.0 [4.8 - 10.0]	8.0 [6.0 - 12.0]	6.0 [4.0 - 9.0]	9.0 [7.0 - 14.0]	11.0 [7.5 - 13.5]	10.0 [7.0 - 14.0]	7.5 [5.0 - 10.0]	<0.001
FVC (% predicted)	97.0 [83.0 - 107.0]	56.0 [56.0 - 56.0]	89.0 [81.0 - 101.0]	92.0 [77.8 - 104.0]	100.0 [86.0 - 109.0]	93.2 [74.5 - 102.6]	98.3 [85.0 - 114.2]	101.0 [92.5 - 109.0]	98.3 [86.5 - 110.8]	102.0 [97.0 - 114.0]	<0.001
Syncope at onset: yes (n [%])	28 [38.4%]	7 [87.5%]	65 [38.2%]	6 [11.3%]	27 [25.2%]	15 [30.6%]	7 [25.0%]	5 [21.7%]	24 [21.6%]	5 [27.8%]	<0.001
RBBB: yes (n [%])	15 [24.6%]	0 [0.0%]	53 [81.5%]	7 [18.9%]	10 [25.6%]	9 [32.1%]	4 [22.2%]	2 [8.7%]	11 [22.9%]	2 [5.7%]	<0.001
Ethnicity: Other (n [%])	0 [0.0%]	1 [12.5%]	67 [28.9%]	1 [1.9%]	9 [7.9%]	12 [22.6%]	9 [31.0%]	1 [4.3%]	7 [6.1%]	23 [53.5%]	<0.001
Height (cm)	163.0 [156.6 - 171.8]	121.2 [107.8 - 135.8]	165.0 [159.0 - 171.0]	162.5 [155.8 - 168.0]	163.0 [158.0 - 170.8]	164.0 [162.0 - 171.0]	165.0 [158.0 - 172.0]	163.0 [156.5 - 173.0]	165.0 [160.0 - 172.0]	169.0 [164.0 - 174.5]	0.001
Vasodilator study CO	3.8 [2.9 - 4.7]	3.1 [2.4 - 3.7]	4.6 [3.2 - 5.3]	4.6 [4.2 - 4.8]	4.7 [3.5 - 5.5]	4.6 [3.0 - 5.3]	5.5 [5.0 - 6.4]	7.3 [7.0 - 7.9]	4.9 [4.1 - 5.5]	5.2 [3.9 - 6.5]	0.001
FVC (L)	3.0 [2.5 - 3.6]	1.9 [1.9 - 1.9]	2.9 [2.3 - 3.6]	2.8 [2.3 - 3.6]	3.0 [2.6 - 4.1]	3.1 [2.7 - 4.0]	3.1 [2.7 - 3.9]	3.0 [2.2 - 3.7]	3.2 [2.7 - 4.0]	3.9 [3.4 - 4.5]	0.001
Systemic SBP (mmHg)	121.0 [110.0 - 130.0]	93.5 [90.8 - 111.2]	126.0 [115.0 - 140.0]	118.0 [110.0 - 125.0]	120.0 [110.0 - 132.5]	117.0 [110.0 - 130.0]	118.0 [110.0 - 131.0]	125.0 [110.5 - 147.0]	124.5 [112.0 - 135.0]	114.0 [110.0 - 120.0]	0.002
MRI RV diastolic volume (ml)	191.0 [191.0 - 191.0]	98.5 [82.8 - 114.2]	203.5 [159.8 - 243.5]	223.0 [210.5 - 235.5]	201.0 [190.0 - 243.0]	190.0 [161.0 - 214.0]	197.0 [121.0 - 262.0]	119.5 [102.2 - 152.2]	146.7 [122.8 - 197.9]	142.0 [120.0 - 160.0]	0.002
sPAP (mmHg)	87.0 [72.0 - 102.0]	46.5 [43.2 - 52.8]	85.0 [71.0 - 99.0]	81.0 [68.0 - 92.0]	88.0 [74.0 - 97.8]	79.0 [64.5 - 101.0]	91.0 [72.0 - 103.5]	91.0 [71.5 - 100.5]	87.0 [76.5 - 105.0]	78.0 [69.0 - 95.0]	0.003
Vasodilator study S_aO_2	98.0 [97.0 - 99.0]	99.2 [99.1 - 99.6]	98.0 [93.8 - 99.0]	99.0 [97.8 - 100.0]	96.0 [94.0 - 98.0]	94.0 [89.0 - 98.0]	94.5 [92.2 - 95.8]	95.0 [92.5 - 96.8]	97.0 [94.5 - 99.0]	97.5 [95.5 - 99.0]	0.003
Dominant R wave: yes (n [%])	30 [48.4%]	1 [33.3%]	86 [72.3%]	15 [40.5%]	11 [40.7%]	7 [33.3%]	10 [45.5%]	13 [56.5%]	15 [39.5%]	20 [57.1%]	0.004

WBC (x10^9/l)	8.0 [6.8 - 9.5]	8.5 [6.9 - 9.8]	8.1 [6.6 - 9.6]	8.9 [6.6 - 9.6]	7.9 [6.6 - 9.3]	9.5 [8.1 - 11.2]	8.8 [7.4 - 11.1]	9.2 [8.2 - 11.0]	8.0 [6.9 - 9.8]	7.9 [6.0 - 9.6]	0.005
Bilirubin (µmol/l)	14.0 [9.2 - 22.8]	13.0 [11.0 - 18.0]	17.0 [11.0 - 26.0]	11.0 [8.0 - 17.0]	17.0 [11.0 - 24.2]	16.0 [10.5 - 20.0]	15.0 [9.0 - 26.2]	12.0 [8.0 - 19.0]	16.0 [10.0 - 22.0]	12.0 [9.0 - 16.5]	0.005
RHC S_aO_2 (%)	96.0 [93.0 - 97.0]	99.0 [99.0 - 99.0]	94.0 [91.0 - 97.0]	95.0 [92.0 - 96.0]	95.0 [92.0 - 97.0]	94.5 [90.8 - 96.0]	96.0 [93.0 - 97.0]	92.0 [90.2 - 95.5]	95.0 [93.0 - 97.0]	96.0 [94.0 - 97.0]	0.005
FEV_1 (L)	2.2 [1.9 - 2.8]	1.6 [1.6 - 1.6]	2.2 [1.8 - 2.7]	2.1 [1.5 - 2.6]	2.4 [1.8 - 3.1]	2.5 [1.7 - 3.2]	2.4 [1.9 - 2.8]	2.1 [1.8 - 2.8]	2.6 [1.9 - 3.0]	2.9 [2.4 - 3.5]	0.005
US portal venous flow: abnormal (n [%])	2 [40.0%]	0 [0.0%]	9 [5.5%]	1 [100.0%]	0 [0.0%]	1 [4.3%]	1 [8.3%]	0 [0.0%]	0 [0.0%]	0 [0.0%]	0.006
RHC DBP (mmHg)	76.0 [71.0 - 85.8]	58.5 [53.5 - 61.2]	74.0 [66.2 - 83.8]	74.5 [66.0 - 81.5]	80.0 [71.0 - 88.0]	75.0 [69.5 - 81.5]	83.0 [71.0 - 88.0]	79.0 [73.5 - 84.0]	78.0 [71.0 - 86.0]	82.0 [69.0 - 89.0]	0.009
ALT (iu/l)	21.0 [16.0 - 29.0]	31.0 [23.0 - 38.0]	23.0 [17.0 - 35.0]	23.5 [17.0 - 31.2]	28.0 [22.0 - 38.0]	23.0 [16.5 - 33.0]	28.0 [21.8 - 36.2]	25.0 [20.5 - 38.0]	22.0 [18.5 - 30.5]	26.0 [20.0 - 40.0]	0.010
MRI RVEF (%)	39.0 [34.5 - 44.4]	58.0 [58.0 - 58.0]	35.0 [25.0 - 43.0]	34.5 [28.8 - 40.2]	36.0 [20.0 - 36.8]	45.0 [40.0 - 51.0]	33.0 [27.9 - 38.5]	46.0 [44.5 - 47.0]	37.0 [24.5 - 42.0]	40.0 [33.0 - 50.0]	0.012
BSA	1.9 [1.7 - 2.0]	0.8 [0.7 - 1.2]	1.9 [1.7 - 2.0]	1.8 [1.6 - 2.0]	1.8 [1.7 - 2.0]	1.8 [1.7 - 2.0]	1.8 [1.7 - 1.9]	1.9 [1.8 - 2.1]	1.9 [1.7 - 2.0]	1.8 [1.7 - 1.9]	0.014
RV on echocardiogram: abnormal (n [%])	70 [95.9%]	8 [100.0%]	218 [97.3%]	47 [97.9%]	96 [94.1%]	46 [93.9%]	21 [77.8%]	21 [91.3%]	74 [90.2%]	18 [94.7%]	0.015
FEV_1 (% predicted)	87.0 [74.0 - 96.0]	52.0 [52.0 - 52.0]	82.1 [72.0 - 93.0]	80.5 [70.2 - 97.0]	88.0 [77.0 - 100.0]	81.2 [67.8 - 90.8]	80.3 [75.4 - 98.5]	90.0 [80.5 - 97.5]	87.5 [77.8 - 98.0]	96.0 [86.0 - 104.0]	0.017
Heart rhythm: sinus / atrial fibrillation / atrial flutter / other (n [%])	61 [95.3%] / 1 [1.6%] / 0 [0.0%] / 2 [3.1%]	5 [100.0%] / 0 [0.0%] / 0 [0.0%] / 0 [0.0%]	211 [97.7%] / 5 [2.3%] / 0 [0.0%] / 0 [0.0%]	38 [100.0%] / 0 [0.0%] / 0 [0.0%] / 0 [0.0%]	44 [95.7%] / 2 [4.3%] / 0 [0.0%] / 0 [0.0%]	39 [100.0%] / 0 [0.0%] / 0 [0.0%] / 0 [0.0%]	24 [100.0%] / 0 [0.0%] / 0 [0.0%] / 0 [0.0%]	20 [87.0%] / 1 [4.3%] / 2 [8.7%] / 0 [0.0%]	75 [97.4%] / 1 [1.3%] / 1 [1.3%] / 0 [0.0%]	37 [97.4%] / 0 [0.0%] / 1 [2.6%] / 0 [0.0%]	0.017
Resting S_aO_2 (%)	94.5 [92.0 - 97.0]	98.0 [98.0 - 99.0]	95.0 [91.0 - 97.0]	94.0 [92.0 - 96.0]	96.0 [92.0 - 97.0]	95.0 [93.0 - 97.0]	95.0 [92.2 - 97.0]	95.0 [91.5 - 97.0]	96.0 [93.0 - 98.0]	97.0 [95.0 - 98.0]	0.028
S_vO_2 (%)	62.9 [58.0 - 68.8]	69.3 [64.2 - 73.2]	64.0 [58.0 - 70.8]	65.0 [60.0 - 71.0]	66.5 [58.8 - 71.8]	64.7 [59.0 - 72.0]	66.0 [60.5 - 70.0]	61.9 [56.5 - 65.5]	62.4 [55.8 - 67.9]	67.5 [59.0 - 72.0]	0.028
Clubbing: yes (n [%])	4 [21.1%]	0 [0.0%]	7 [11.7%]	1 [2.0%]	8 [7.9%]	0 [0.0%]	2 [7.4%]	0 [0.0%]	1 [1.7%]	0 [0.0%]	0.028
Ankle swelling: yes (n [%])	33 [45.8%]	0 [0.0%]	66 [36.9%]	16 [32.0%]	32 [30.5%]	9 [22.0%]	2 [7.1%]	9 [39.1%]	32 [29.1%]	8 [22.9%]	0.028
Raynauds: yes (n [%])	11 [22.9%]	0 [0.0%]	12 [9.1%]	3 [5.9%]	6 [5.6%]	1 [2.3%]	1 [3.6%]	1 [4.3%]	7 [6.1%]	1 [7.1%]	0.028
Potassium (mmol/l)	4.3 [4.0 - 4.6]	3.9 [3.9 - 4.1]	4.2 [4.0 - 4.4]	4.3 [4.1 - 4.6]	4.2 [3.9 - 4.5]	4.2 [3.9 - 4.4]	4.4 [4.3 - 4.7]	4.2 [3.9 - 4.4]	4.2 [4.0 - 4.5]	4.2 [3.9 - 4.3]	0.031

Systemic DBP (mmHg)	74.0 [66.5 - 81.0]	57.5 [52.5 - 62.5]	78.5 [68.0 - 86.0]	72.0 [65.0 - 80.0]	76.0 [70.0 - 83.0]	74.0 [66.0 - 80.0]	78.0 [70.0 - 85.0]	82.0 [71.0 - 87.0]	78.5 [68.0 - 85.0]	80.0 [73.5 - 83.8]	0.031
Infection at onset: yes (n [%])	7 [33.3%]	0 [0.0%]	17 [17.5%]	6 [12.0%]	6 [5.6%]	3 [6.7%]	2 [7.1%]	2 [9.1%]	6 [12.2%]	1 [6.2%]	0.031
Pre-walk S_aO_2 (%)	94.5 [91.0 - 97.0]	95.0 [95.0 - 95.0]	96.0 [93.0 - 97.0]	94.0 [91.0 - 96.8]	96.0 [93.0 - 98.0]	97.0 [94.5 - 98.0]	96.5 [94.0 - 97.0]	96.0 [95.0 - 97.0]	95.0 [92.0 - 97.0]	97.0 [93.2 - 98.0]	0.032
ANA: positive (n [%])	10 [18.9%]	0 [0.0%]	21 [11.9%]	6 [15.0%]	3 [4.3%]	1 [5.0%]	3 [13.6%]	4 [18.2%]	17 [22.4%]	8 [30.8%]	0.039
CRP (mg/l)	5.5 [3.0 - 11.0]	5.0 [5.0 - 6.0]	4.0 [2.0 - 8.0]	5.0 [5.0 - 6.5]	5.0 [2.0 - 7.2]	5.0 [2.0 - 9.0]	3.0 [1.0 - 4.0]	9.0 [3.0 - 11.0]	3.9 [1.9 - 8.6]	5.3 [3.6 - 6.7]	0.040
TAPSE (cm)	1.4 [1.2 - 1.9]	1.9 [1.6 - 2.1]	1.5 [1.2 - 1.9]	1.6 [1.4 - 2.0]	1.6 [1.3 - 1.9]	1.6 [1.3 - 1.7]	2.0 [1.8 - 2.2]	1.6 [1.3 - 2.0]	1.4 [1.2 - 1.8]	1.5 [1.2 - 1.8]	0.041

mPAP – mean pulmonary artery pressure, PCWP – pulmonary capillary wedge pressure, CO – cardiac output, CI – cardiac index, BMI – body mass index, ALP – alkaline phosphatase, 6mwt – six minute walk test, dPAP – diastolic pulmonary artery pressure, RAP – right atrial pressure, FVC – forced vital capacity, RBBB – right bundle branch block, SBP – systolic blood pressure, RV – right ventricle, sPAP – systolic pulmonary artery pressure, WBC – white blood cells, RHC – right heart catheterisation, FEV_1 – forced expiratory volume in 1 second, US – ultrasound, DBP – diastolic blood pressure, ALT – alanine transaminase, MRI – magnetic resonance imaging, RVEF – right ventricular ejection fraction, BSA – body surface area, S_vO_2 – mixed venous oxygen saturation, ANA – anti-nuclear antibodies, CRP – C reactive protein, TAPSE – tricuspid annular plane systolic excursion. Data presented as median [IQR] unless stated. Kruskal-Wallis test used to assess continuous variables, Fisher's exact test used to assess categorical variables, Cochran–Armitage test used to assess ordinal variables.

Appendix 13

Ethnicity codes adapted from the NHS data dictionary:

White

A British

B Irish

C Any other White background

Mixed

D White and Black Caribbean

E White and Black African

F White and Asian

G Any other mixed background

Asian or Asian British

H Indian

J Pakistani

K Bangladeshi

L Any other Asian background

Black or Black British

M Caribbean

N African

P Any other Black background

Other Ethnic Groups

R Chinese

S Any other ethnic group

Z Not stated

Appendix 14

Variables of prognostic significance in univariate Cox proportional hazards models							
Variable	**Categorical group**	**Number of patients**	**Number of events**	**Hazard ratio**	**Confidence interval**	**Corrected p**	**Left truncated analysis: corrected p**
Age at diagnosis		*655*	*155*	*1.049*	*1.038 - 1.060*	*<0.001*	*<0.001*
P_aO_2		231	57	0.811	0.727 - 0.903	0.025	
Red cell distribution width		*200*	*46*	*1.216*	*1.102 - 1.341*	*0.015*	*0.041*
WBC		487	118	1.137	1.059 - 1.221	0.068	
Albumin		465	117	0.938	0.905 - 0.973	0.079	
Urea		*482*	*118*	*1.141*	*1.098 - 1.187*	*<0.001*	*0.002*
Estimated GFR		250	61	0.975	0.962 - 0.988	0.037	
S_aO_2 at rest		*449*	*110*	*0.926*	*0.900 - 0.952*	*<0.001*	*<0.001*
S_aO_2 pre walk test		*437*	*100*	*0.914*	*0.885 - 0.944*	*<0.001*	*<0.001*
S_aO_2 post walk test		*394*	*93*	*0.960*	*0.940 - 0.981*	*0.036*	*0.034*
Functional class		621	149	1.743	1.345 - 2.258	0.005	
RAP		577	135	1.065	1.035 - 1.096	0.003	
S_aO_2 during RHC		*439*	*93*	*0.908*	*0.880 - 0.937*	*<0.001*	*0.002*
S_vO_2		432	95	0.932	0.912 - 0.953	<0.001	
KCO		*377*	*81*	*0.075*	*0.040 - 0.142*	*<0.001*	*<0.001*
KCO % predicted		*366*	*84*	*0.956*	*0.946 - 0.967*	*<0.001*	*<0.001*
KCO z score		53	13	0.272	0.135 - 0.548	0.043	
6mwt distance		*203*	*43*	*0.996*	*0.994 - 0.998*	*0.029*	*0.002*
Gender	*Male*	*655*	*155*	*1.919*	*1.398 - 2.634*	*0.010*	*0.029*
Ascites	Present	368	73	6.465	2.943 - 14.198	<0.001	
RBBB	Present	256	55	3.025	1.769 - 5.170	0.008	
Incident / Prevalent	Prevalent group	655	155	0.367	0.234 - 0.576	0.001	
Age group	Young group	655	155	0.283	0.201 - 0.397	<0.001	
Rhythm	AF	408	97	4.793	2.085 - 11.016	0.035	

AF – atrial fibrillation, GFR – glomerular filtration rate, KCO – transfer coefficient for carbon monoxide, P_aO_2 – Partial pressure of oxygen in arterial blood, RAP – right atrial pressure, RBBB – right bundle branch block, S_aO_2 – peripheral arterial oxygen saturation, WBC – white blood cells. *Italic* – significant in left truncated analysis.

Appendix 15

Phenotypic differences between male and female patients with idiopathic PAH			
	Female	**Male**	**Corrected p**
n [%]	500 [68.6%]	229 [31.4%]	
WGS population: African / East-Asian / European / South-Asian (n [%])	17 [3.4%] / 12 [2.4%] / 439 [87.8%] / 32 [6.4%]	5 [2.2%] / 0 [0.0%] / 210 [91.7%] / 14 [6.1%]	0.190
Gender: female (n [%])	500 [100.0%]	0 [0.0%]	<0.001
Age at diagnosis (years)	48.8 [34.7 - 62.0]	59.5 [42.3 - 68.5]	<0.001
mPAP (mmHg)	53.0 [44.0 - 62.0]	49.0 [41.0 - 56.0]	0.010
PCWP (mmHg)	9.0 [7.0 - 11.0]	10.0 [7.0 - 12.0]	0.672
CO (L/min)	3.9 [3.2 - 4.9]	4.4 [3.5 - 5.7]	<0.001
CI (L/min/m^2)	2.2 [1.8 - 2.7]	2.1 [1.8 - 2.7]	0.952
PVR (WU)	11.1 [7.4 - 14.8]	8.5 [6.3 - 12.5]	<0.001
Vasoresponder: (n [%])	22 [18.8%]	8 [17.8%]	1.000
Functional class: 1 / 2 / 3 / 4 (n [%])	8 [1.9%] / 91 [21.8%] / 271 [64.8%] / 48 [11.5%]	7 [3.4%] / 46 [22.5%] / 131 [64.2%] / 20 [9.8%]	0.540
6mwt distance (m)	305.0 [153.0 - 396.8]	325.0 [212.2 - 426.8]	0.382
Height (cm)	162.0 [157.0 - 166.0]	174.0 [167.5 - 179.0]	<0.001
Weight (kg)	71.4 [61.0 - 85.2]	84.0 [73.0 - 95.5]	<0.001
BSA	1.8 [1.6 - 1.9]	2.0 [1.9 - 2.1]	<0.001
Platelets (x10^9/l)	242.0 [198.5 - 294.0]	197.0 [169.0 - 242.0]	<0.001
HCT	0.4 [0.4 - 0.5]	0.5 [0.4 - 0.5]	<0.001
Hb (g/l)	147.5 [134.0 - 159.0]	159.0 [146.0 - 169.0]	<0.001
Ferritin (μg/l)	64.0 [28.0 - 109.0]	121.0 [59.5 - 243.5]	<0.001
Urea (mmol/l)	5.4 [4.2 - 6.9]	6.8 [5.2 - 9.4]	<0.001
Creatinine (μmol/l)	81.0 [70.0 - 95.0]	104.0 [86.0 - 123.0]	<0.001
MRI LV diastolic volume (ml)	80.0 [66.6 - 105.0]	101.0 [91.0 - 143.0]	<0.001
FEV_1 (L)	2.0 [1.6 - 2.5]	2.7 [2.2 - 3.2]	<0.001

FVC (L)	2.8 [2.2 - 3.3]	3.9 [3.2 - 4.5]	<0.001
TLC (L)	4.6 [4.0 - 5.1]	6.2 [5.5 - 6.7]	<0.001
KCO (mmol/min/kPa/l)	1.2 [0.8 - 1.4]	0.9 [0.6 - 1.3]	<0.001
VA (L)	4.0 [3.4 - 4.6]	5.5 [5.0 - 6.3]	<0.001
Coronary artery disease: Yes (n [%])	16 [4.4%]	32 [19.4%]	<0.001
Post walk S_aO_2 (%)	93.0 [86.0 - 96.0]	88.0 [82.0 - 94.0]	0.004
MRI LV systolic volume (ml)	28.2 [20.0 - 43.2]	40.5 [30.8 - 68.9]	0.004
Potassium (mmol/l)	4.2 [3.9 - 4.4]	4.3 [4.1 - 4.6]	0.006
Resting S_aO_2 (%)	96.0 [92.0 - 98.0]	94.0 [91.0 - 97.0]	0.008
Bilirubin (μmol/l)	14.0 [9.0 - 22.0]	15.5 [12.0 - 23.0]	0.011
Transferrin (g/l)	3.0 [2.6 - 3.4]	2.5 [2.2 - 3.0]	0.014
Pre walk S_aO_2 (%)	96.0 [93.0 - 98.0]	95.0 [90.2 - 97.0]	0.014
CT emphysema: none / minimal / mild / moderate / severe (n [%])	225 [94.5%] / 3 [1.3%] / 7 [2.9%] / 3 [1.3%] / 0 [0.0%]	98 [83.1%] / 6 [5.1%] / 5 [4.2%] / 8 [6.8%] / 1 [0.8%]	0.014
CPET height (m)	164.5 [157.0 - 170.0]	175.5 [167.0 - 178.2]	0.015
LA size (cm)	3.5 [2.9 - 3.8]	3.9 [3.4 - 4.1]	0.017
Obesity: Yes (n [%])	10 [2.8%]	15 [9.1%]	0.017
sPAP (mmHg)	86.0 [71.0 - 100.0]	79.0 [68.0 - 93.0]	0.018
RHC S_aO_2 (%)	95.0 [92.0 - 97.0]	94.0 [91.0 - 96.0]	0.018
Total cholesterol (mmol/l)	4.7 [3.9 - 5.5]	4.2 [3.5 - 4.8]	0.024
CPET Peak PET CO_2	3.1 [2.6 - 3.6]	2.0 [1.8 - 2.5]	0.025
P_aO_2 (kPa)	9.3 [7.5 - 11.1]	7.9 [7.1 - 9.8]	0.030
TLC (% predicted)	97.0 [86.0 - 104.0]	91.0 [83.0 - 99.0]	0.032
RHC HR (bpm)	79.0 [70.0 - 90.0]	74.0 [65.0 - 86.0]	0.035
CT fibrosis: none / minimal / mild / moderate / severe (n [%])	231 [96.2%] / 7 [2.9%] / 2 [0.8%] / 0 [0.0%] / 0 [0.0%]	105 [89.0%] / 5 [4.2%] / 6 [5.1%] / 1 [0.8%] / 1 [0.8%]	0.049

6mwt – six minute walk test, BSA – body surface area, CI – cardiac index, CO – cardiac output, CPET – cardiopulmonary exercise test, CT – computer tomography, FEV_1 – forced expiratory volume in 1 second, FVC – forced vital capacity, Hb – haemoglobin, HCT – haematocrit, HR – heart rate, KCO – transfer coefficient for carbon monoxide, LA – left atrium, LV – left ventricle, WGS – whole genome sequencing, mPAP – mean pulmonary artery pressure, MRI – magnetic resonance imaging, PCWP – pulmonary capillary wedge pressure, PET CO_2 – partial pressure of end tidal carbon dioxide, PVR – pulmonary vascular resistance, RHC – right heart catheterisation, sPAP – systolic pulmonary artery pressure, TLC – total lung capacity, VA – alveolar volume. Data presented as median [IQR] unless

stated. Wilcoxon rank-sum test used to assess continuous variables, Fisher's exact test used to assess categorical variables, Cochran–Armitage test used to assess ordinal variables.

Phenotypic differences between young and old patients with idiopathic PAH			
	Young	**Old**	**Corrected p**
n [%]	363 [50.0%]	363 [50.0%]	
WGS population: African / East-Asian / European / South-Asian (n [%])	10 [2.8%] / 10 [2.8%] / 310 [85.4%] / 33 [9.1%]	12 [3.3%] / 2 [0.6%] / 336 [92.6%] / 13 [3.6%]	0.004
Gender: female (n [%])	276 [76.0%]	221 [60.9%]	<0.001
Age at diagnosis (years)	38.1 [28.5 - 44.8]	65.7 [58.6 - 71.7]	<0.001
mPAP (mmHg)	54.0 [46.0 - 63.0]	49.0 [41.0 - 57.5]	<0.001
PCWP (mmHg)	9.0 [6.0 - 11.0]	10.0 [7.8 - 12.0]	<0.001
CI (L/min/m^2)	2.3 [1.9 - 2.9]	2.1 [1.7 - 2.6]	0.006
PVR (WU)	10.7 [7.0 - 15.3]	10.1 [7.0 - 13.4]	0.055
Vasoresponder: (n [%])	24 [24.2%]	6 [9.5%]	0.059
Functional class: 1 / 2 / 3 / 4 (n [%])	12 [3.8%] / 95 [30.2%] / 173 [54.9%] / 35 [11.1%]	3 [1.0%] / 42 [13.7%] / 229 [74.8%] / 32 [10.5%]	<0.001
6mwt distance (m)	380.0 [300.0 - 438.0]	217.5 [96.0 - 330.0]	<0.001
P_aO_2 (kPa)	9.4 [7.9 - 11.6]	8.3 [7.0 - 9.8]	<0.001
BMI	26.4 [22.1 - 31.2]	28.7 [25.2 - 32.8]	<0.001
Urate (mmol/l)	0.4 [0.3 - 0.4]	0.5 [0.4 - 0.6]	<0.001
ALP (iu)	78.0 [65.0 - 104.0]	93.0 [76.0 - 117.5]	<0.001
Urea (mmol/l)	4.8 [3.9 - 5.8]	6.9 [5.4 - 9.3]	<0.001
Creatinine (μmol/l)	79.0 [69.0 - 94.0]	95.0 [80.0 - 113.0]	<0.001
eGFR (ml/min)	79.0 [67.0 - 89.2]	58.2 [45.5 - 70.8]	<0.001
Resting S_aO_2 (%)	97.0 [95.0 - 98.0]	93.0 [90.2 - 96.8]	<0.001
Systemic SBP (mmHg)	118.0 [110.0 - 130.0]	129.5 [115.0 - 142.8]	<0.001
Pre walk S_aO_2 (%)	97.0 [95.0 - 98.0]	94.0 [90.0 - 96.0]	<0.001
Post walk S_aO_2 (%)	94.0 [88.0 - 97.0]	88.0 [82.0 - 93.0]	<0.001
dPAP (mmHg)	35.0 [29.0 - 41.2]	30.0 [24.0 - 36.0]	<0.001
S_vO_2 (%)	67.8 [61.6 - 72.8]	63.5 [58.0 - 69.5]	<0.001
FEV_1 (L)	2.6 [2.1 - 3.1]	2.0 [1.6 - 2.5]	<0.001
FVC (L)	3.3 [2.8 - 3.9]	2.8 [2.1 - 3.5]	<0.001
KCO (mmol/min/kPa/l)	1.4 [1.2 - 1.6]	0.8 [0.6 - 1.1]	<0.001

KCO (% predicted)	79.0 [69.0 - 91.0]	55.8 [40.0 - 75.0]	<0.001
Syncope at onset: yes (n [%])	80 [34.9%]	42 [18.3%]	<0.001
Lung function pattern: Normal / Obstructive / Restrictive (n [%])	126 [66.0%] / 24 [12.6%] / 41 [21.5%]	100 [42.7%] / 102 [43.6%] / 32 [13.7%]	<0.001
Diabetes mellitus: Yes (n [%])	9 [3.5%]	41 [15.5%]	<0.001
Coronary artery disease: Yes (n [%])	1 [0.4%]	47 [17.8%]	<0.001
Systemic hypertension: Yes (n [%])	5 [1.9%]	31 [11.7%]	<0.001
Hyperlipidaemia: Yes (n [%])	2 [0.8%]	28 [10.6%]	<0.001
Congenital heart disease: Yes (n [%])	26 [10.0%]	6 [2.3%]	<0.001
COPD or ILD: Yes (n [%])	5 [1.9%]	41 [15.5%]	<0.001
Red cell distribution width	13.9 [12.9 - 15.5]	14.7 [13.9 - 16.6]	0.002
Alveolar volume (% predicted)	89.5 [82.0 - 100.0]	85.0 [77.0 - 93.2]	0.004
Ankle swelling: Yes (n [%])	58 [25.2%]	93 [40.4%]	0.004
CT emphysema: none / minimal / mild / moderate / severe (n [%])	157 [96.9%] / 3 [1.9%] / 0 [0.0%] / 2 [1.2%] / 0 [0.0%]	166 [85.6%] / 6 [3.1%] / 12 [6.2%] / 9 [4.6%] / 1 [0.5%]	0.004
FVC (% predicted)	92.0 [80.0 - 103.0]	98.0 [84.8 - 111.0]	0.005
NT-ProBNP (ng/l)	567.6 [140.9 - 1463.0]	1564.5 [359.7 - 3181.8]	0.006
HR (bpm)	80.0 [72.0 - 93.0]	80.0 [70.0 - 86.0]	0.006
Weight (kg)	73.0 [59.7 - 88.0]	78.8 [67.5 - 90.5]	0.007
Dominant R wave: Yes (n [%])	80 [61.1%]	59 [41.5%]	0.007
CT fibrosis: none / minimal / mild / moderate / severe (n [%])	160 [98.2%] / 2 [1.2%] / 0 [0.0%] / 1 [0.6%] / 0 [0.0%]	176 [90.3%] / 10 [5.1%] / 8 [4.1%] / 0 [0.0%] / 1 [0.5%]	0.007

RA area (cm^2)	20.1 [16.0 - 24.6]	23.4 [19.9 - 30.0]	0.009
Ascites: Yes (n [%])	1 [0.5%]	10 [5.5%]	0.016
ANCA: Positive (n [%])	4 [4.2%]	18 [16.2%]	0.019
CPET Peak PET CO_2	3.3 [2.8 - 3.7]	2.4 [2.1 - 3.1]	0.038
Height (cm)	165.0 [160.0 - 172.0]	164.0 [158.0 - 170.2]	0.044
Valvular heart disease: Yes (n [%])	0 [0.0%]	7 [2.7%]	0.044
CPET VO_2-WR slope	8.3 [6.3 - 9.3]	6.1 [4.0 - 7.3]	0.045
Transferrin (g/l)	3.0 [2.7 - 3.4]	2.8 [2.5 - 3.2]	0.046

6mwt – six minute walk test, ALP – alkaline phosphatase, ANCA – anti-neutrophil cytoplasmic antibodies, BMI – body mass index, CI – cardiac index, COPD – chronic obstructive pulmonary disease, CPET – cardiopulmonary exercise test, CT – computer tomography, dPAP – diastolic pulmonary artery pressure, eGFR – estimated glomerular filtration rate, FEV_1 – forced expiratory volume in 1 second, FVC – forced vital capacity, HR – heart rate, ILD – interstitial lung disease, KCO – transfer coefficient for carbon monoxide, mPAP – mean pulmonary artery pressure, NT-ProBNP – N terminal pro bone natriuretic peptide, PCWP – pulmonary capillary wedge pressure, PET CO_2 – partial pressure of end tidal carbon dioxide, PVR – pulmonary vascular resistance, RA – right atrium, SBP – systolic blood pressure, WGS – whole genome sequencing, WR – work rate. Data presented as median [IQR] unless stated. Wilcoxon rank-sum test used to assess continuous variables, Fisher's exact test used to assess categorical variables, Cochran–Armitage test used to assess ordinal variables.

Phenotypic differences between idiopathic PAH patients with high and low K_{CO} % predicted			
	High	**Low**	**Corrected p**
n [%]	180 [49.2%]	186 [50.8%]	
WGS population: African / East-Asian / European / South-Asian (n [%])	4 [2.2%] / 2 [1.1%] / 160 [88.9%] / 14 [7.8%]	4 [2.2%] / 2 [1.1%] / 169 [90.9%] / 11 [5.9%]	0.999
Gender: Female (n [%])	130 [72.2%]	118 [63.4%]	0.254
Age at diagnosis (years)	45.2 [35.3 - 57.4]	64.7 [51.6 - 71.7]	<0.001
mPAP (mmHg)	55.0 [46.0 - 64.5]	49.0 [42.0 - 56.0]	<0.001
PCWP (mmHg)	10.0 [7.0 - 12.0]	10.0 [8.0 - 12.0]	0.850
CI (L/min/m^2)	2.1 [1.7 - 2.7]	2.2 [1.8 - 2.5]	0.903
PVR (WU)	11.3 [7.4 - 15.4]	10.1 [7.1 - 13.4]	0.367
Vasoresponder: (n [%])	14 [25.5%]	8 [14.8%]	0.536
Functional class: 1 / 2 / 3 / 4 (n [%])	3 [1.7%] / 28 [16.2%] / 121 [69.9%] / 21 [12.1%]	0 [0.0%] / 24 [13.5%] / 129 [72.5%] / 25 [14.0%]	0.457
6mwt distance (m)	330.0 [260.0 - 408.0]	232.0 [96.0 - 395.0]	0.028
Urea (mmol/l)	5.4 [4.3 - 6.7]	6.4 [5.1 - 8.3]	<0.001
Resting S_aO_2 (%)	96.0 [94.0 - 98.0]	93.0 [90.0 - 97.0]	<0.001
Pre walk S_aO_2 (%)	96.0 [94.0 - 98.0]	94.0 [90.0 - 97.0]	<0.001
Post walk S_aO_2 (%)	94.0 [89.0 - 96.0]	87.0 [80.8 - 93.0]	<0.001
K_{CO} (mmol/min/kPa/l)	1.4 [1.3 - 1.6]	0.7 [0.5 - 0.9]	<0.001
K_{CO} (% predicted)	84.5 [77.9 - 94.2]	48.4 [38.0 - 61.4]	<0.001
CT emphysema: none / minimal / mild / moderate / severe (n [%])	126 [98.4%] / 1 [0.8%] / 1 [0.8%] / 0 [0.0%] / 0 [0.0%]	110 [82.7%] / 6 [4.5%] / 10 [7.5%] / 6 [4.5%] / 1 [0.8%]	<0.001
Lung function pattern: Normal / Obstructive / Restrictive (n [%])	107 [60.8%] / 32 [18.2%] / 37 [21.0%]	90 [48.9%] / 77 [41.8%] / 17 [9.2%]	<0.001
COPD or ILD: Yes (n [%])	5 [2.8%]	34 [18.5%]	<0.001

dPAP (mmHg)	35.0 [28.0 - 40.0]	30.0 [25.0 - 36.0]	0.001
FVC (% predicted)	91.0 [81.0 - 103.0]	100.0 [86.8 - 111.0]	0.004
P_aO_2 (kPa)	9.3 [7.7 - 11.5]	8.2 [6.6 - 9.8]	0.009
sPAP (mmHg)	90.0 [74.0 - 102.0]	80.0 [69.0 - 94.0]	0.009
Iron (μmol/l)	16.4 [11.6 - 21.0]	12.9 [8.0 - 17.0]	0.011
S_vO_2 (%)	67.7 [60.8 - 72.7]	63.7 [58.0 - 69.2]	0.011
Coronary artery disease: Yes (n [%])	9 [5.1%]	29 [15.8%]	0.011
eGFR (ml/min)	72.0 [59.5 - 83.0]	60.0 [50.0 - 73.2]	0.022
FEV_1 (L)	2.4 [1.9 - 3.0]	2.2 [1.6 - 2.7]	0.023
CPET VE-VCO_2 slope	39.8 [32.3 - 56.2]	77.0 [51.4 - 82.5]	0.030
Hb (g/l)	155.0 [139.5 - 168.5]	150.0 [133.0 - 160.0]	0.032
CPET Peak O_2 pulse	8.1 [6.1 - 9.1]	6.3 [5.0 - 6.9]	0.032
Congenital heart disease: Yes (n [%])	20 [11.2%]	6 [3.3%]	0.032
CPET OUES	1113.0 [919.0 - 1274.0]	632.5 [403.0 - 851.8]	0.035
RDW	14.2 [13.2 - 15.1]	14.5 [13.9 - 16.5]	0.039
Systemic DBP (mmHg)	80.0 [70.0 - 89.0]	76.0 [67.0 - 83.0]	0.043
CPET Peak PET CO_2	3.3 [2.7 - 3.6]	2.3 [2.0 - 2.9]	0.043

6mwt – six minute walk test, CI – cardiac index, COPD – chronic obstructive pulmonary disease, CPET – cardiopulmonary exercise test, CT – computer tomography, DBP – diastolic blood pressure, dPAP – diastolic pulmonary artery pressure, eGFR – estimated glomerular filtration rate, FEV_1 – forced expiratory volume in 1 second, FVC – forced vital capacity, Hb – haemoglobin, ILD – interstitial lung disease, K_{CO} – transfer coefficient for carbon monoxide, mPAP – mean pulmonary artery pressure, OUES – oxygen uptake efficiency slope, PCWP – pulmonary capillary wedge pressure, PET CO_2 – partial pressure of end tidal carbon dioxide, PVR – pulmonary vascular resistance, RDW – red cell distribution width, sPAP – systolic pulmonary artery pressure, VE – expired ventilation, VE-VCO_2 - ventilatory equivalent for carbon dioxide, WGS – whole genome sequencing. Data presented as median [IQR] unless stated. Wilcoxon rank-sum test used to assess continuous variables, Fisher's exact test used to assess categorical variables, Cochran–Armitage test used to assess ordinal variables.

Appendix 16

Rare and predicted deleterious *BMPR2* variants identified in PAH patients							
HGVSc	**HGVSp**	**CSQ**	**Domain**	**Gender**	**Age**	**FHx**	**Other variant**
c.19delC	p.Arg7GlyfsTer40	Frameshift	ECD	Female	29	No	
c.26_41delGGCGGGTGCCCTGGCT	p.Trp9TyrfsTer33	Frameshift	ECD	Female	39	No	
c.27G>A	p.Trp9Ter	Stop gained	ECD	Female	27	Yes	
c.48G>A	p.Trp16Ter	Stop gained	ECD	Female	50	Yes	
c.76+1G>A		Splice region		Female	29	No	
c.76+2T>G		Splice region		Male	51	No	
c.93_94insT	p.Arg32SerfsTer6	Frameshift	ECD	Male	48	No	
c.100T>C	p.Cys34Arg	Missense	ECD	Female	42	No	
c.100T>C	p.Cys34Arg	Missense	ECD	Female	30	Yes	
c.156_157delTC	p.His53Ter	Frameshift	ECD	Male	20	Yes	
c.156_157delTC	p.His53Ter	Frameshift	ECD	Male	41	Yes	*SMAD1* (c.400+1G>A)
c.156_157delTC	p.His53Ter	Frameshift	ECD	Female	58	No	
c.200A>G	p.Tyr67Cys	Missense	ECD	Female	54	Yes	
c.203G>A	p.Gly68Asp	Missense	ECD	Female	59	No	
c.218C>G	p.Ser73Ter	Stop gained	ECD	Female	73	No	
c.244C>T	p.Gln82Ter	Stop gained	ECD	Female	40	No	

c.247+1G>A		Splice region		Female	54	No	
c.251G>A	p.Cys84Tyr	Missense	ECD	Female	42	Yes	
c.274delC	p.Gln92LysfsTer9	Frameshift	ECD	Female	27	No	
c.280T>C	p.Cys94Arg	Missense	ECD	Male	44	No	
c.280T>G	p.Cys94Gly	Missense	ECD	Male	37	No	
c.288T>G	p.Tyr96Ter	Stop gained	ECD	Female	40	No	
c.314delC	p.Pro105LeufsTer47	Frameshift	ECD	Male	64	No	
c.344dupT	p.Cys116LeufsTer4	Frameshift	ECD	Male	30	No	
c.346T>C	p.Cys116Arg	Missense	ECD	Female	33	Yes	
c.346T>C	p.Cys116Arg	Missense	ECD	Male	35	Yes	
c.349T>G	p.Cys117Gly	Missense	ECD	Male	28	No	
c.367T>C	p.Cys123Arg	Missense	ECD	Male	38	No	*SMAD9* (p.Arg294Ter)
c.367T>C	p.Cys123Arg	Missense	ECD	Female	38	No	
c.371dupA	p.Asn124LysfsTer6	Frameshift	ECD	Female	23	Yes	
c.377A>G	p.Asn126Ser	Missense	ECD	Male	52	Yes	
c.399delT	p.Pro134LeufsTer18	Frameshift	ECD	Female	34	Yes	
c.408_412delAACAC	p.Pro138GlnfsTer7	Frameshift	ECD	Female	39	No	
c.417_418+2delCAGT		Splice region	ECD	Female	42	No	
c.418+1G>T		Splice region		Male	47	Yes	
c.418+1G>T		Splice region		Female	13	Yes	

c.418+5G>A		Splice region		Male	44	No	
c.439C>T	p.Arg147Ter	Stop gained	TM	Female	65	No	
c.439C>T	p.Arg147Ter	Stop gained	TM	Female	55	No	
c.439C>T	p.Arg147Ter	Stop gained	TM	Female	40	No	
c.449dupC	p.Ile151AsnfsTer30	Frameshift	TM	Female	37	No	
c.470C>G	p.Ser157Ter	Stop gained	TM	Female	58	No	
c.529+2dupT		Splice region		Female	23	Yes	
c.529+2dupT		Splice region		Female	46	Yes	
c.533_536dupACCG	p.Lys180ProfsTer2	Frameshift	PK	Male	34	No	
c.612delA	p.Lys204AsnfsTer5	Frameshift	PK	Male	45	Yes	
c.619dupG	p.Glu207GlyfsTer13	Frameshift	PK	Female	45	No	
c.621+1G>A		Splice region		Female	35	No	
c.621+1G>C		Splice region		Female	19	No	
c.621+1G>C		Splice region		Female	66	Yes	
c.621+1G>C		Splice region		Male	43	No	
c.622-1G>T		Splice region		Male	44	No	
c.631C>T	p.Arg211Ter	Stop gained	PK	Female	32	No	
c.637C>T	p.Arg213Ter	Stop gained	PK	Female	28	No	
c.657delA	p.Gly220AlafsTer10	Frameshift	PK	Female	64	No	
c.657delA	p.Gly220AlafsTer10	Frameshift	PK	Female	62	No	
c.683delC	p.Ala228ValfsTer2	Frameshift	PK	Female	47	No	

c.691delG	p.Val231CysfsTer21	Frameshift	PK	Male	79	No	
c.761_762delAT	p.His254ArgfsTer11	Frameshift	PK	Male	29	No	
c.761_762delAT	p.His254ArgfsTer11	Frameshift	PK	Female		No	
c.793G>T	p.Glu265Ter	Stop gained	PK	Male	51	No	
c.796_799delAGAG	p.Arg266SerfsTer12	Frameshift	PK	Female	51	No	
c.823dupT	p.Tyr275LeufsTer23	Frameshift	PK	Female	34	No	
c.843C>G	p.Tyr281Ter	Stop gained	PK	Female	37	No	
c.852_852+1insA	p.Gly285ArgfsTer13	Frameshift	PK	Female	28	No	
c.853-1G>A		Splice region		Female	47	No	
c.853-2A>G		Splice region		Male	44	No	
c.860T>A	p.Leu287Ter	Stop gained	PK	Male	34	Yes	
c.917A>C	p.His306Pro	Missense	PK	Female	62	No	
c.961C>T	p.Arg321Ter	Stop gained	PK	Male	45	Yes	
c.961C>T	p.Arg321Ter	Stop gained	PK	Female	43	No	
c.967+4delA		Splice region		Female	32	No	
c.967+4delA		Splice region		Female	54	Yes	
c.968-1G>T		Splice region		Female	39	No	
c.992A>G	p.His331Arg	Missense	PK	Female	34	No	
c.994C>T	p.Arg332Ter	Stop gained	PK	Male	59	Yes	
c.994C>T	p.Arg332Ter	Stop gained	PK	Male	38	No	
c.994C>T	p.Arg332Ter	Stop gained	PK	Female	66	No	

c.994C>T	p.Arg332Ter	Stop gained	PK	Female	49	No	
c.994C>T	p.Arg332Ter	Stop gained	PK	Female	42	No	
c.1019T>C	p.Leu340Pro	Missense	PK	Female	55	Yes	
c.1039T>C	p.Cys347Arg	Missense	PK	Male	20	Yes	
c.1128+2T>G		Splice region		Female	36	Yes	
c.1128+2T>G		Splice region		Female	53	No	
c.1133G>T	p.Gly378Val	Missense	PK	Male	34	No	
c.1178A>G	p.Asn393Ser	Missense	PK	Female	24	No	
c.1202T>C	p.Leu401Ser	Missense	PK	Male	42	No	
c.1202T>C	p.Leu401Ser	Missense	PK	Female	31	Yes	
c.1217T>G	p.Met406Arg	Missense	PK	Male	52	Yes	
c.1221T>G	p.Tyr407Ter	Stop gained	PK	Female	27	Yes	
c.1228G>C	p.Gly410Arg	Missense	PK	Female	59	No	
c.1242G>A	p.Trp414Ter	Stop gained	PK	Male	22	Yes	
c.1245_1246dupGA	p.Ile416ArgfsTer4	Frameshift	PK	Male	39	Yes	
c.1255_1257dupAGA	p.Arg419dup	Inframe indel	PK	Female	32	No	
c.1277-2A>G		Splice region		Male	35	No	
c.1355_1356dupTC	p.Val453SerfsTer22	Frameshift	PK	Female	61	No	
c.1371dupA	p.Gln458ThrfsTer13	Frameshift	PK	Male	60	Yes	
c.1398G>A	p.Trp466Ter	Stop gained	PK	Male	67	No	
c.1413+1G>A		Splice region		Female	42	No	

c.1424C>A	p.Ser475Ter	Stop gained	PK	Female	63	Yes	
c.1432G>T	p.Glu478Ter	Stop gained	PK	Male	31	No	
c.1454A>G	p.Asp485Gly	Missense	PK	Female	30	No	
c.1471C>T	p.Arg491Trp	Missense	PK	Female	27	No	
c.1471C>T	p.Arg491Trp	Missense	PK	Female	22	Yes	*BMPR2* (p.Arg591Gln)
c.1471C>T	p.Arg491Trp	Missense	PK	Female	30	Yes	*BMPR2* (p.Arg591Gln)
c.1471C>T	p.Arg491Trp	Missense	PK	Female	38	No	
c.1471C>T	p.Arg491Trp	Missense	PK	Female	30	Yes	
c.1471C>T	p.Arg491Trp	Missense	PK	Male	56	Yes	
c.1472G>A	p.Arg491Gln	Missense	PK	Female	49	No	
c.1472G>A	p.Arg491Gln	Missense	PK	Male	29	No	
c.1472G>A	p.Arg491Gln	Missense	PK	Female	35	Yes	
c.1490_1507delCTGAGGAAAGGATGGCTG	p.Ala497_Ala502del	Inframe indel	PK	Male	36	Yes	
c.1606C>T	p.Arg536Cys	Missense	CT	Male	71	No	
c.1750C>T	p.Arg584Ter	Stop gained	CT	Female	73	No	
c.1865C>A	p.Pro622Gln	Missense	CT	Female	51	No	
c.1958_1959delCT	p.Pro653ArgfsTer21	Frameshift	CT	Female	59	No	
c.1962_1963insGA	p.Cys655AspfsTer5	Frameshift	CT	Male	39	No	
c.2004delA	p.Asp669ThrfsTer31	Frameshift	CT	Female	74	No	

c.2014G>T	p.Glu672Ter	Stop gained	CT	Male	51	No	
c.2027_2030dupACCT	p.Lys678ProfsTer6	Frameshift	CT	Female	51	No	*SMAD9* (p.Gly367Ser)
c.2158C>T	p.Gln720Ter	Stop gained	CT	Male	38	No	
c.2245dupC	p.Gln749ProfsTer9	Frameshift	CT	Female	30	No	
c.2372_2373delTG	p.Met791LysfsTer21	Frameshift	CT	Male	39	Yes	
c.2426dupT	p.Ala810GlyfsTer3	Frameshift	CT	Female	27	No	
c.2500C>T	p.Gln834Ter	Stop gained	CT	Female	42	No	
c.2533delG	p.Glu845LysfsTer14	Frameshift	CT	Female	24	No	
c.2542C>T	p.Gln848Ter	Stop gained	CT	Male	45	No	
c.2548C>T	p.Gln850Ter	Stop gained	CT	Male	45	Yes	
c.2558_2559insA	p.Glu854Ter	Frameshift	CT	Female	39	No	
c.2608_2612delTTACT	p.Leu870GlufsTer9	Frameshift	CT	Male	37	No	
c.2617C>T	p.Arg873Ter	Stop gained	CT	Female	37	No	
c.2617C>T	p.Arg873Ter	Stop gained	CT	Female	19	No	
c.2617C>T	p.Arg873Ter	Stop gained	CT	Male	59	No	
c.2617C>T	p.Arg873Ter	Stop gained	CT	Female	82	Yes	
c.2695C>T	p.Arg899Ter	Stop gained	CT	Female	32	No	
c.2695C>T	p.Arg899Ter	Stop gained	CT	Female	48	No	*EIF2AK4* (p.Arg899Ter)
c.2866+448_*10129del		Deletion		Female	37	Yes	
c.-2135886_*1477184del		Deletion		Female	36	No	

c.418+6917_968-2529del		Deletion		Male	44	No	
c.1413+4060_1586+1016del		Deletion		Female	32	No	*ENG* (p.Cys652Tyr)
c.77-4215_418+117del		Deletion		Female	30	No	
c.-469639_*1793991del		Deletion		Female	43	No	
c.-18799_76+6577del		Deletion		Female	38	No	
c.-14078_*480624del		Deletion		Female	27	No	
c.-8079_76+173del		Deletion		Male	37	No	
c.818_967+3133del		Deletion		Female	56	Yes	*ENG* (p.Thr617Met)
c.1277-3948_1414-2306del		Deletion		Female	27	No	
c.247+28_418+3004del		Deletion		Female	53	No	
c.530-336_967+4272del		Deletion		Female	27	No	
c.-1420039_*2025475del		Deletion		Male	29	No	
c.967+1209_*20667del		Deletion		Female	34	No	
c.1277-573_1414-2475del		Deletion		Female	61	No	
c.77-2315_248-646del		Deletion		Female	38	Yes	
c.1277-942_1414-4738del		Deletion		Female	36	Yes	
c.77-1125_418+99del		Deletion		Male	46	No	
c.418+9084_2866+804del		Deletion		Female	32	No	

c.-243014_*61340del		Deletion		Female	64	No	

CSQ – consequence type, CT – cytoplasmic tail, ECD – extracellular domain, FHx – family history of PAH, PK – protein kinase domain, TM – transmembrane domain.

Appendix 17

Phenotypic differences between PAH patients with and without *BMPR2* variants			
	***BMPR2*$^{+/-}$**	***BMPR2*$^{+/+}$**	**Corrected p**
n [%]	157 [16.8%]	779 [83.2%]	
WGS population: African / East-Asian / European / South-Asian (n [%])	3 [1.9%] / 2 [1.3%] / 147 [93.6%] / 5 [3.2%]	24 [3.1%] / 14 [1.8%] / 693 [89.0%] / 48 [6.2%]	0.741
Gender: female (n [%])	105 [66.9%]	544 [69.9%]	0.751
Age at diagnosis (years)	39.3 [32.0 - 51.1]	51.4 [38.2 - 65.5]	<0.001
Drug exposure: Yes (n [%])	6 [3.8%]	49 [6.3%]	0.523
mPAP (mmHg)	58.0 [51.0 - 68.2]	51.0 [43.0 - 60.0]	<0.001
PCWP (mmHg)	10.0 [7.0 - 12.0]	9.0 [7.0 - 11.0]	0.819
CO (L/min)	3.3 [2.7 - 4.0]	4.1 [3.3 - 5.2]	<0.001
CI (L/min/m^2)	1.8 [1.5 - 2.2]	2.2 [1.8 - 2.7]	<0.001
PVR (WU)	14.2 [10.8 - 20.3]	10.2 [7.0 - 14.0]	<0.001
Vasoresponder (n [%])	0 [0.0%]	31 [18.1%]	0.007
Functional class: 1 / 2 / 3 / 4 (n [%])	2 [1.4%] / 32 [21.6%] / 89 [60.1%] / 25 [16.9%]	15 [2.2%] / 143 [21.4%] / 439 [65.7%] / 71 [10.6%]	0.388
6mwt distance (m)	331.0 [286.5 - 426.5]	310.0 [160.0 - 410.0]	0.21
HCT	0.5 [0.5 - 0.5]	0.4 [0.4 - 0.5]	<0.001
Hb (g/l)	162.0 [152.0 - 173.2]	151.0 [136.0 - 163.0]	<0.001
ALT (iu/l)	30.0 [22.0 - 42.0]	23.0 [17.0 - 33.0]	<0.001
sPAP (mmHg)	92.0 [82.0 - 105.0]	84.0 [70.0 - 97.0]	<0.001
dPAP (mmHg)	40.0 [33.2 - 47.8]	33.0 [26.0 - 39.0]	<0.001
S_vO_2 (%)	60.1 [56.0 - 65.9]	65.4 [59.6 - 71.3]	<0.001
FEV_1 (L)	2.6 [2.2 - 3.3]	2.2 [1.8 - 2.8]	<0.001
FVC (L)	3.5 [2.9 - 4.2]	3.0 [2.3 - 3.8]	<0.001
KCO (mmol/min/kPa/litre)	1.4 [1.2 - 1.6]	1.1 [0.7 - 1.4]	<0.001

K_{CO} (% predicted)	82.0 [74.0 - 94.0]	70.2 [49.0 - 85.4]	<0.001
Lung function pattern: Normal / Obstructive / Restrictive (n [%])	81 [81.0%] / 12 [12.0%] / 7 [7.0%]	244 [53.7%] / 135 [29.7%] / 75 [16.5%]	<0.001
Bilirubin (µmol/l)	19.0 [12.5 - 29.0]	14.0 [10.0 - 22.0]	0.003
TSH (mu/l)	2.5 [1.7 - 3.7]	2.0 [1.2 - 3.1]	0.013
HR (bpm)	85.0 [73.0 - 96.0]	78.0 [69.0 - 88.0]	0.013
WBC ($x10^9$)	8.8 [7.3 - 10.8]	8.1 [6.8 - 9.7]	0.014
Post walk S_aO_2 (%)	94.0 [89.8 - 97.0]	91.0 [85.0 - 96.0]	0.014
FEV_1 (% predicted)	90.8 [78.0 - 99.4]	85.4 [73.0 - 95.5]	0.033
Coronary artery disease: Yes (n [%])	2 [1.7%]	49 [8.8%]	0.034

6mwt – six minute walk test, ALT – alanine transferase, CI – cardiac index, CO – cardiac output, dPAP – diastolic pulmonary artery pressure, FEV_1 – forced expiratory volume in 1 second, FVC – forced vital capacity, Hb – haemoglobin, HCT – haematocrit, HR – heart rate, K_{CO} – transfer coefficient for carbon monoxide, mPAP – mean pulmonary artery pressure, PCWP – pulmonary capillary wedge pressure, PVR – pulmonary vascular resistance, sPAP – systolic pulmonary artery pressure, TSH – thyroid stimulating hormone, WGS – whole genome sequencing. Data presented as median [IQR] unless stated. Wilcoxon rank-sum test used to assess continuous variables, Fisher's exact test used to assess categorical variables, Cochran–Armitage test used to assess ordinal variables.

Appendix 18

Rare and predicted deleterious *TBX4* variants identified in PAH patients							
HGVSc	**HGVSp**	**CSQ**	**SPS**	**Gender**	**Age**	**FHx**	**Other variant**
c.40_49delTTCCGGGCCC	p.Phe14Argfs Ter28	Frameshift		Female	30	No	
c.121G>T	p.Gly41Ter	Stop gained		Male	56	No	
c.884G>C	p.Gly295Ala	Missense		Male	10	Yes	
c.885dupC	p.Thr296Hisfs Ter90	Frameshift		Female	74	No	
c.972delT	p.Thr326Profs Ter12	Frameshift		Female	79	Yes	*ENG* (p.Gly645 Ala)
c.1001_1004dupATCA	p.His335Glnfs Ter52	Frameshift		Male	5	No	
c.1012A>T	p.Lys338Ter	Stop gained		Male	29	No	
c.1070C>T	p.Ala357Val	Missense		Female	81	No	
c.1102C>T	p.Arg368Cys	Missense		Female	39	No	
c.1112dupC	p.Pro372Serfs Ter14	Frameshift	Yes	Male	57	No	
c.1112dupC	p.Pro372Serfs Ter14	Frameshift	Yes	Male	74	No	
c.1145A>C	p.Tyr382Ser	Missense		Female	67	No	
c.1160_1167delCCCCCAGA	p.Thr387Argfs Ter29	Frameshift		Female	57	No	
c.1183T>C	p.Ser395Pro	Missense		Female	33	No	
c.1274C>A	p.Pro425Gln	Missense		Female	49	No	
c.1543G>A	p.Glu515Lys	Missense		Male	41	No	
CSQ – consequence type, FHx – family history of PAH, SPS – previously associated with Small Patella Syndrome.							

Appendix 19

Rare and predicted deleterious *KCNK3* variants identified in PAH patients					
HGVSc	**HGVSp**	**Consequence type**	**Gender**	**Age at diagnosis**	**Family history**
c.152G>T	p.Arg51Leu	Missense	Female	73	No
c.340G>A	p.Ala114Thr	Missense	Male	27	No
c.544G>C	p.Glu182Gln	Missense	Male	21	No
c.707G>C	p.Gly236Ala	Missense	Female	27	No

Rare and predicted deleterious *SMAD9* variants identified in PAH patients						
HGVSc	**HGVSp**	**CSQ**	**Gender**	**Age at diagnosis**	**FHx**	**Other variants**
c.295C>G	p.Pro99Ala	Missense	Male	51	No	
c.296C>T	p.Pro99Leu	Missense	Female	61	No	
c.790C>A	p.Pro264Thr	Missense	Female	28	No	
c.880C>T	p.Arg294Ter	Stop gained	Male	38	No	*BMPR2* (p.C123R)
c.1099G>A	p.Gly367Ser	Missense	Female	51	No	*BMPR2* (p.Lys678Profs Ter6)
CSQ – consequence type, FHx – family history of PAH.						

Appendix 20

Rare and predicted deleterious *EIF2AK4* variants identified									
HGVSc	**HGVSp**	**CSQ**	**Genotype**	**Gender**	**Age**	**FHx**	**Diagnosis**	**Kco % predicted**	**Other**
c.1072_1073dupGT	p.Val359Ter	Frameshift	Heterozygote	Female	39	No	PAH	72	
c.44C>T	p.Pro15Leu	Missense	Heterozygote	Female	40	No	PAH	109	
c.3722A>G	p.Glu1241Gly	Missense	Heterozygote	Female	25	No	PAH	41	
c.1660G>T	p.Asp554Tyr	Missense & splice region	Heterozygote	Male	72	No	PAH		
c.3711_3713delGAG	p.Arg1238del	Inframe deletion	Heterozygote	Female	59	No	PAH	95	
c.3604C>T	p.His1202Tyr	Missense	Heterozygote	Female	48	No	PAH	61	*BMPR2* (p.Arg899Ter)
c.220G>A	p.Asp74Asn	Missense	Heterozygote	Female	70	No	PAH		
c.2446C>T ^	p.Gln816Ter	Stop gained	Heterozygote	Female	24	No	PAH	81	
c.3218G>T ^	p.Arg1073Leu	Missense	Heterozygote						

c.3605A>T	p.His1202Leu	Missense	Homozygote	Female	36	No	PAH	40	
c.1795G>C	p.Gly599Arg	Missense	Homozygote	Male	22	No	PAH	31	
c.4392dupT	p.Lys1465Ter	Frameshift & splice region	Homozygote	Male	21	No	PVOD / PCH	33	
c.3097C>T	p.Gln1033Ter	Stop gained	Homozygote	Male	29	Yes	PAH	27	
c.1159_1160delCT	p.Leu387CysfsTer27	Frameshift	Homozygote	Male	18	No	PAH	28	
c.1795G>C	p.Gly599Arg	Missense	Homozygote	Female	25	No	PAH	33	
c.281dupA	p.Asn94LysfsTer8	Frameshift	Homozygote	Male	39	No	PVOD / PCH		
c.3055_3064delCTGACCAACG	p.Leu1019TrpfsTer9	Frameshift	Potential compound heterozygote	Male	23	No	PAH	33	
c.3884T>G	p.Leu1295Arg	Missense							
c.4400dupT	p.Glu1468ArgfsTer14	Frameshift	Potential compound heterozygote	Male	48	No	PAH	45	
c.1739dupA	p.Arg581GlufsTer9	Frameshift							
c.1392delT	p.Arg465ValfsTer38	Frameshift	Potential compound heterozygote	Female	70	No	PAH	33	*SMAD9* (p.Pro140Leu)
c.257+4A>C		Splice region							

c.1820T>G	p.Val607Gly	Missense & splice region	Potential compound heterozygote	Female	27	No	PVOD / PCH	28	*SMAD9* (p.Gly367 Ser)
c.2727C>G	p.Ser909Arg	Missense							
c.3325G>A	p.Gly1109Arg	Missense	Potential compound heterozygote	Female	33	No	PVOD / PCH		
c.2841delG	p.Ile948SerfsTer35	Frameshift							
c.145-2A>G		Splice acceptor	Potential compound heterozygote	Female	38	No	PAH		
c.4418_4421delCAGA	p.Thr1473ArgfsTer17	Frameshift							
c.2827A>G	p.Thr943Ala	Missense							
c.3097C>T	p.Gln1033Ter	Stop gained	Potential compound heterozygote	Female	50	Yes	PVOD / PCH	30	
c.4769delT	p.Leu1590Ter	Frameshift							
c.551_552delAA *	p.Glu184GlyfsTer13	Frameshift	Potential compound heterozygote	Male	43	No	PVOD / PCH	46	
c.2721T>G *	p.Tyr907Ter	Stop gained							

^ – also identified in mother, * – identified using the modified variant filtering strategy, CSQ – consequence type, FHx – family history of PAH or PVOD / PCH, K_{CO} – transfer coefficient for carbon monoxide